Exploring the contribution of voice quality to the perception of gender in Scottish English

Jo Pearce

Submitted in fulfilment of the requirements for the Degree of Master of Philosophy

School of Critical Studies

College of Arts

University of Glasgow

June 2020

Exploring the contribution of voice quality to the perception of gender in Scottish English

Jo Pearce

Abstract

This study investigates how voice quality, here phonation, affects listener perception of speaker gender, and how voice quality interacts with pitch, a major cue to speaker gender, when cueing gender perceptions. Gender differences in voice quality have been identified in both Scottish (Beck and Schaeffler 2015; Stuart-Smith 1999) and American English (Abdelli-Beruh et al. 2014; D. Klatt and L. Klatt 1990; Podesva 2013; Syrdal 1996; Wolk et al. 2012; Yuasa 2010). There is evidence from previous research that suggest gender differences in voice quality may also influence listener perception of speaker gender, with breathy voice being perceived as feminine or female characteristic by listeners (Addington 1968; Andrews and Schmidt 1997; Bishop and Keating 2012; Holmberg et al. 2010; Porter 2012; Skuk and Schweinberger 2014; Van Borsel et al. 2009) and creaky voice being perceived as masculine characteristic (Greer 2015; Lee 2016). However, some studies have found that voice quality has little effect (Booz and Ferguson 2016; King et al. 2012; Owen and Hancock 2010). The present study seeks to investigate the contribution of voice quality, taking into account the various methods of producing voice quality differences in stimuli, cultural differences in gendered meanings of voice quality, and different methods of quantifying 'perceived gender', which may contribute to the conflicting results of previous studies.

To investigate the contribution of voice quality to perceptions of speaker gender, a perception experiment was be carried out where 32 Scottish listeners and 40 North American listeners heard stimuli with different voice qualities (modal, breathy, creaky) and at different pitch levels (120Hz, 165Hz, 210Hz), and were asked to make judgements about the gender of the speaker. Differences in voice quality were produced by a speaker with the ability to create voice quality distinctions, as well as created through copy synthesis from the speaker's voice. Listeners were asked to indicate whether they thought the voice belonged to a man or a woman and rate how masculine and feminine the voice sounded. Relative to modal voice, I predicted that listeners would be more likely to categorise breathy voices as women, and would rate them as more feminine and less masculine, and that listeners would be less likely to categorise creaky voices as women, and would rate them as more masculine and less feminine. I also predicted that there might be differences in how Scottish listeners and North American listeners perceived voice quality, given that the gender differences in voice quality in these two varieties of English have been found to differ in previous research.

Consistent with my predictions, I found that relative to modal voice, listeners were more likely to categorise breathy voice stimuli as women, and rated breathy voice stimuli as more feminine and less masculine. However, in contrast with my predictions, I found that relative to modal voice, listeners were more likely to categorise creaky voice stimuli as women, and rated them as less masculine, but not more feminine. Furthermore, contrary to predictions, I did not identify differences between Scottish and North American listeners in terms of voice quality perception. Differences were also found in how breathy and creaky voice influence gender perception at different pitch levels.

Overall, these results show that voice quality has an important influence on listener perception of speaker gender, and that the gendered meanings of creaky voice are changing and have disassociated from its low pitch. Future research should consider whether this evaluation among Scottish listeners this may reflect a wider change in the gender differences in production.

Acknowledgments

First of all, I would like to say a big thank thank you to both of my supervisors, Clara Cohen and Jane Stuart-Smith, not only for your input and support on this project over the past year, which has been so valuable, but also for both being incredible lecturers and researchers in my undergraduate years and inspiring me to take this on in the first place.

I'd also like to thank my participants, who made this project possible by letting me use their voices and taking part in my experiment (and an extra thank you to those who gave up their time to let me record them, but who I didn't end up using in the final experiment).

A huge thank you as well to everyone in GULP who has helped me out over the past year, and taken many, many coffee breaks with me.

Ethical clearance for this project has been granted by the College of Arts Research Ethics committee (07/02/2019).

This work was supported by the Economic and Social Research Council.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This study investigates how voice quality, here phonation, affects listener perception of speaker gender, and how voice quality interacts with pitch, a major cue to speaker gender, when cueing gender perceptions. While it has often been suggested that voice quality affects listener perceptions of speaker gender (Addington 1968; Andrews and Schmidt 1997; Bishop and Keating 2012; Holmberg et al. 2010; Porter 2012; Skuk and Schweinberger 2014; Van Borsel et al. 2009), the results of previous studies have produced conflicting results (Booz and Ferguson 2016; King et al. 2012; Owen and Hancock 2010) and the role of voice quality in gender perception is still unknown. The term 'gender' here will be used in a broad sense, encompassing listeners' perception of a voice as belonging to a man or a woman, as well as judgements of masculinity and femininity, whatever these terms may mean to each listener.

Understanding how voice quality affects listener perceptions of speaker gender is important for several reasons. Firstly, perception is integral in understanding differences in production; authors often speculate about why they find gender differences in voice quality production, and link these variably to physiological causes, indexical meanings of voice quality, or potential iconic links between voice quality and gender. Investigating perception of voice quality can help to interrogate these claims. Furthermore, it can help us to understand how gender non-conformity is evaluated in voices. Finally, voice quality changes are often a goal in voice therapy for transgender individuals; examining how voice quality contributes to gender perception therefore allows evaluation of the suitability of these goals.

To investigate the contribution of voice quality to perceptions of speaker gender, a perception experiment was carried out where listeners heard stimuli with either modal, breathy or creaky voice quality, and pitch that was manipulated to fall in either an androgynous, typical male or typical female level. Listeners were asked to indicate whether they thought the voice belonged to a man or a woman and rate how masculine and feminine the voice sounded.

In Chapter 1, I give an overview of the research background and introduce the topic of the present study. I begin in Section 1.1 with an overview of the research context that this study is situated in, giving an overview of gender differences in speech and important cues to speaker gender in the voice. In Section 1.2, I focus in on voice quality as a cue of particular interest, as its role in gender perception has been debated in previous research. I outline how differences in voice quality are produced in the vocal tract, the ways of measuring voice quality differences acoustically, and the gender difference that have been found in voice quality production in different varieties of English. In Section 1.3, I turn to the role of voice quality in the perception of gender and explore the findings of previous research on this topic. This overview of the research background finishes in Section 1.4 with a summary of my research questions and predictions.

In Chapter 2, I detail the methods of my two experiments, one conducted in-person with Scottish listeners, and one conducted online with North American listeners. In these experiments, listeners listened to stimuli with naturally produced and synthesized voice quality differences, and made judgements about the gender of the voice they heard. Section 2.1 describes the methods I used to create my stimuli and create voice quality differences in them, using synthesis in KlattGrid and asking a speaker to produce differences naturally. Section 2.2 looks at the participants, procedure and design of Experiment 1, conducted in-person with Scottish listeners. Section 2.3 looks at the participants, procedure and design of Experiment 2, conducted online with North American listeners. Section 2.4 describes the statistical analysis of my results. In Section 2.5, I detail the qualitative methods I took to analysis participant feedback in my experiments.

In Chapter 3, I present my results, starting in Section 3.1 with listeners' ratings of femininity, then in Section 3.2 looking at listeners' ratings of masculinity, and in Section 3.2 looking at listeners' categorisations of gender. To preview the findings, compared to modal voice, breathy voice is perceived as more feminine, less masculine and more likely to be the voice of a woman, while creaky voice is perceived as less masculine and more likely to be the voice of a woman, but not more feminine.

In Chapter 4, I first discuss results first individually by the effect of each independent variable that I considered in Section 4.1, and then in 4.2 consider the results together and discuss them in light of my research questions and their wider theoretical implications. In Section 4.3, I discuss the methodology of the study.

Finally, Chapter 5, I wrap up with some concluding remarks.

1.1 Doing gender and hearing gender

1.1.1 Conceptualising ‘sex’ and ‘gender’

As has been noted in both sociological (e.g. Bradley 2014; Jackson and Scott 2001) and linguistic (e.g. Bucholtz 2002; Kiesling 2019; McElhinny 2014) work on gender, the concepts of sex and gender are difficult to define and understandings of them have changed drastically in recent years. In order to investigate how listeners perceive gender in the voice, we must first explore the concepts of ‘sex’ and ‘gender’, as well as the relationship between them.

In biologically deterministic views of sex and gender, the term ‘gender’, if used at all, is seen as equivalent to ‘sex’, and the two sexes of ‘male’ and ‘female’ are seen as easy to separate by chromosomes and morphological differences in the body. In the 1970s, social and feminist theorists working on gender began to separate the terms sex and gender, talking of sex as biological, but gender as a social, cultural, or psychological term (e.g. Oakley 2016[1972]; Rubin 1975). However, as West and Zimmerman (1987) note, this distinction becomes hard to maintain because of the complex relationship between biological and cultural processes. They propose an alternate distinction between *sex*, *sex category* and *gender*. They define sex as the agreed-upon set of biological criteria, such as chromosomes and genitalia, that are used to classify people into a sex category, either male or female, which stands in as a proxy for sex in everyday life, but can vary independently from it. This is a useful conceptualisation within phonetics, as sex differences that tend to exist in the morphology of the vocal tract, such as the shape and size of the larynx, can lead to broad differences in speech between speakers of male and female sex categories. Gender, in turn, is anything we *do* in interaction to proclaim our membership of a sex category; it is the activity of managing our behaviour in light of societal expectations of what is appropriate for our sex category (West and Zimmerman 1987).

In a similar vein to West and Zimmerman (1987), Butler (2014[1990]) interrogates the idea that gender is simply the cultural counterpart of sex, and argues that sex categories themselves are gendered and may in fact be just as much of a cultural construct as gender itself. She then proposes that gender is performative and that it is produced through stylised, repeated performance of mundane acts; by doing certain things, such as acting or dressing, in particular ways, we do not reveal an essential gender identity, but instead bring gender itself into being.

One criticism of performativity is that it ignores the importance of the body in performances and personal experiences of gender; Butler (2014[1993]) addresses this concern and argues that just because gender is an abstract, constructed system, this does not mean that it does not interact with the physical body. This idea lays the foundation for discussions of gender embodiment, a term that refers to how gender — as an individual identity, as a product of social interactions, and as a component of social institutions — interacts with experiences of living within particular bodies (Mason 2018). Taking the evidence from studies on transgender speech, Zimman (2018) demonstrates the usefulness of embodiment when considering gender differences in speech and the voice, as these are often a product of the interaction between physiological differences and gender performance, even where physiology has often assumed to be the sole cause of differences. He argues that the use of sex and gender terms like 'female voice' and 'male voice' is problematized by these insights from studies on transgender speech, and that carefully defining and distinguishing between the relevant aspects of sex and gender is vitally important when studying gender and the voice.

Here, I will use terms like 'male voice' and 'female voice' when discussing previous research as they are used in the study that I am referring to; generally, as Zimman (2018) points out, this is as a shorthand for 'the voice of a person deemed to be (fe)male', either by the researchers' assumptions or as the participants have identified themselves. I will also use terms like 'sex characteristics' and 'sexual dimorphism' in line with West and Zimmerman (1987)'s use of 'sex', to mean the socially agreed-upon biological criteria, such as laryngeal growth or genitalia, that are used to classify individuals into various categories. As I am focusing on the perception of voices, I will be making use of the term 'perceived gender' to refer broadly to the how listeners perceive a voice as male or female, masculine or feminine, or as belonging to a man or a woman, whatever these terms may mean to each listener. I will also use the more specific term 'perceived gender attributes' to refer to listeners' perception of the masculinity and femininity of a voice. It is important to note that while some listeners may take these terms to be similar, others may take them to mean very different things.

1.1.2 Indexical and iconic signals of gender in speech

One of the many ways that we perform gender is through speech. As outlined by Silverstein (1976), the social behaviour of speech, including the performance of gender, can be understood in terms of semiotics. In Peirce et al. (1931)'s theory of semiotics, a sign is anything that conveys meaning that is not the sign itself; in the case of words, the sign is generally *symbolic*, meaning that the connection between the sign and the thing that it signifies is arbitrary. However, sometimes the connection between the sign and what it signifies is not entirely arbitrary. For *iconic* signs, there is some sort of physical or perceivable connection between the sign and what it signifies, like in the case of onomatopoeia. For *indexical* signs, there is some sort of *contextual* link between the sign and the thing it signifies - the sign 'points to' the context that it often occurs in. According to Silverstein (1976), speech consists of not just symbolic signs, but also indexical signs that signal, or *index*, wider social information. Silverstein (2003)'s theory of the indexical order outlines how linguistic features can come to index a certain social situation or group. Indexicality also provides a framework for understanding how gender is performed linguistically (e.g. Ochs 1992). This is demonstrated in work by Eckert (2000, 2001, 2008) and Eckert and McConnell-Ginet (2013) on how gender is constructed linguistically in a Detroit high school.

Eckert (2000) examines the linguistic construction of gender in two locally-situated social groups: middle-class, school-oriented Jocks and working class, urban-oriented Burnouts. She notes that Jocks and Burnouts use different variants of several vowels, but that within this, there are also differences between girls and boys. By using certain vocalic variants, and combining these with other linguistic and non-linguistic resources like clothing style, Jock and Burnout girls and boys perform gender in distinct ways and cause these vocalic variants to become associated with these distinct styles. These vocalic variants can then be said to index the wider linguistic style that they form a part of and 'point to' the Jock or Burnout context that they are associated with. Therefore, similar to the performance of gender more broadly, by not only speaking but also behaving, dressing and so on, in certain ways, the Jock and Burnout adolescents bring Jock and Burnout identity into being, as well as gendered ideas of what it means to be a Jock or Burnout girl or boy.

However, as Eckert (2017) notes, some signs that carry social information may have not only indexical, but also *iconic* meaning; that is, they have some sort of physical or perceivable link with the thing that they signify, rather than just a contextual one. One example of a sign with iconic meaning is pitch.

Pitch tends to differ between adult male and female speakers, with male voices tending to sound lower and female voices tending to sound higher. In spoken English, fundamental frequency (f0), the acoustic correlate of pitch, averages approximately 100-120 Hz in cisgender male speakers and 200-220 Hz in cisgender female speakers (Simpson 2009). A major factor that contributes to this difference is exposure to testosterone during puberty which results in laryngeal changes and a lowering of f0 among male speakers (Hollien et al. 1994). As Hinton et al. (1994) notes in his discussion of the frequency code, higher frequencies are associated with smallness and larger frequencies are associated with largeness, which speakers then exploit in displays of gender. This leads to pitch being an iconic sign of speaker gender, with listeners linking the tendency for male speakers to have a lower f0 with male sex characteristics more generally. Through consideration of how gender differences for f0 vary between cultures, languages, and different varieties of the same languages, we can see that gender differences in f0, though influenced by physiology, are also culturally-specific and may be exaggerated or minimised.

Traunmüller and Eriksson (1995) compare findings of average speaking fundamental frequency across languages note that in stark contrast to findings on English varieties, in Wù dialects of Chinese male speakers average 170 Hz, while female speakers average 187 Hz (Rose 1991). Similarly, Japanese-speaking men have been found to use a lower f0 than English-speaking men, while Japanese-speaking women have been found to use a higher f0 than English-speaking women (Loveday 1981; Yuasa 2008). Even within English, Szakay (2006) finds that New Zealand speakers of Maori English have a significantly higher f0 than speakers of Paheka English, while Szakay and Torgersen (2015) find that pitch differs between speakers from Inner and Outer London as well as between speakers with local or non-local family roots. Furthermore, there is some evidence to suggest that children begin to differentiate their speaking fundamental frequency even before puberty begins (Ferrand and Bloom 1996; Hasek et al. 1980; Ingrisano et al. 1980), although this has not been found in all studies (Busby and Plant 1995).

This goes to show that while there is an iconic meaning to gender differences in pitch, due to laryngeal differences leading to a tendency for male speakers to have lower f0 than female speakers, there is also an indexical meaning: this tendency is then drawn on by speakers to perform their gender identity, creating cross-linguistic differences in pitch within genders that cannot be understood without considering the influence of social factors.

While pitch is one of the most socially salient gender differences, it is by no means the only difference between the speech of cisgender male and female speakers. Simpson (2009) gives an overview of gender differences often found in speech. An important difference is vocal tract resonance, where vowel formant frequencies tend to be higher among female speakers than in male speakers. Although this trend is to be expected from differences in the dimensions of the vocal tract above the glottis, these differences cannot explain the size of formant frequency differences. Simpson (2009) notes that there is still uncertainty as to what explains the size of the difference, but that it is likely to be an interaction of social factors, as the size of the gender difference varies between languages (Henton 1995), and other physiological factors.

Gender differences also exist in phonation, a sub-component of voice quality to do with the vibration of the vocal folds that encompasses settings like modal, breathy, and creaky voice. Gender differences in phonation have been found in various varieties of English in the UK (Beck and Schaeffler 2015; Gittelson et al. 2018; Henton and Bladon 1985, 1988; Stuart-Smith 1999; Szakay and Torgersen 2015), with a general tendency for male speakers to use more creaky voice and female speakers to use more breathy voice. Gender differences in phonation have also been found in varieties of English spoken outside of the UK; in the US in particular, young female speakers are particularly creaky (Abdelli-Beruh et al. 2014; Podesva and Chun 2010; Wolk et al. 2012; Yuasa 2010). There is still ongoing disagreement as to what causes gender differences in voice quality, and whether these differences are entirely indexical, to an extent connected to iconic meanings of pitch, influenced by physiology, or some combination of these potential factors.

1.1.3 Hearing gender

We have seen that there are various differences in speech and voice between male and female speakers; we will now turn to discussion of how listeners perceive these differences, and how listeners perceive gender in the speakers they hear.

Numerous studies looking at various varieties of English, mostly American English, have investigated which cues are most important to listener perception of speaker gender, but they have largely been limited to investigating the relative contribution of fundamental frequency and vowel formant frequencies to listener perception of speaker gender. In varieties of English, differences in f0 between male and female speakers have been found to contribute to listeners' gender perception, with voices with higher f0 more likely to be perceived as a woman/female, and voices with a lower f0 more likely to be perceived as a man/male (Bishop and Keating 2012; Gallena et al. 2018; Gelfer and Mikos 2005; Hillenbrand and Clark 2009; Holmberg et al. 2010; Hubbard and Assmann 2013; King et al. 2012; Lass et al. 1976; Porter 2012; Whiteside 1998; Wolfe et al. 1990). Other studies have found that differences in f0 between voices affect listeners' perception of gender attributes, with lower f0 voices perceived as more masculine, and higher f0 voices perceived as more feminine (Booz and Ferguson 2016; Cartei et al. 2014; Coleman 1976; Gelfer and Schofield 2000; Hardy et al. 2016; Munson 2007; Owen and Hancock 2010).

However, as f0 can overlap between male and female speakers and is not the only gender difference in speech, simply shifting the f0 of a voice does not always successfully alter the perception of its gender. In a systematic review of 38 articles examining the contributions of aspects of verbal communication contributing to listener perceptions of speaker gender across various varieties of English, Leung et al. (2018) found that speaking fundamental frequency accounted for 41.6% of the variance in gender perception, indicating that while f0 is important, a range of other aspects of speech and the voice also contribute to gender perception, to varying degrees. According to Leung et al. (2018)'s review, vowel formant frequencies are the most-studied, and have been found to make an important contribution to gender perception. While most research on this agrees that altering both f0 and vowel formant frequencies together is more likely to change the gender perception of a voice than shifting either of these alone, there is still ongoing debate about the relative contributions of resonance and f0 to gender perception. While some studies find that f0 is more important to gender perception than vowel formant frequencies (Gelfer and Mikos 2005; Lass et al. 1976;

Skuk and Schweinberger 2014; Whiteside 1998), other studies find that vowel formant frequencies may actually be more important than f0 (Gelfer and Bennett 2013; Pisanski and Rendall 2011).

Reviewing the findings of previous work (Hancock et al. 2014; King et al. 2012; Wolfe et al. 1990), Leung et al. (2018) find that there is also evidence that intonation may contribute to the perception of gender. Additionally, there is some limited evidence that loudness (Andrews and Schmidt 1997), vowel space area (Booz and Ferguson 2016), and the articulation and acoustic characteristics of /s/ (Avery and Liss 1996; Munson 2007) may contribute to gender perceptions, but Leung et al. (2018) stress that existing research on these cues is not conclusive and thus further research on these characteristics is needed.

Phonation type may also influence perception of speaker gender. Some studies have found that breathier voices (Addington 1968; Andrews and Schmidt 1997; Van Borsel et al. 2009) or voices that avoid creakiness (Holmberg et al. 2010) are more likely to be rated as female or more feminine by listeners. Furthermore, other studies find that differences in acoustic measures of voice quality affect listener judgements of speaker gender, but do not connect these measures to a perceptual voice quality (Bishop and Keating 2012; Porter 2012; Skuk and Schweinberger 2014). However, other studies find that there is no contribution of voice quality to gender identification (Booz and Ferguson 2016; King et al. 2012; Owen and Hancock 2010). As highlighted by Leung et al. (2018) and Davies et al. (2015), there is a lack of research addressing the contribution of voice quality to the perception of speaker gender and the conflicting results of the existing research demonstrate a need for further studies investigating the contribution of voice quality to perceptions of speaker gender.

1.2 Voice quality, phonation and gender

Laver (1980: 1) defines voice quality as ‘the characteristic auditory colouring of an individual speaker’s voice’. The term ‘voice quality’ encompasses both those aspects that are influenced by physiology, and those that are controlled by the speaker, for example, to indicate their membership of a certain group (Abercrombie 1967: 92). This project will focus on phonation, a subsection of voice quality settings to do with laryngeal configuration, and specifically on breathy and creaky voice, as it is in phonation that most differences are found between male and female speakers (Simpson 2009). Both physiological and social factors may contribute to substantial differences in voice quality often being found between male and female speakers across varieties of English. In order to examine whether, and if so, how listeners use voice quality information to judge speaker gender, we must first understand what voice quality is and how it can

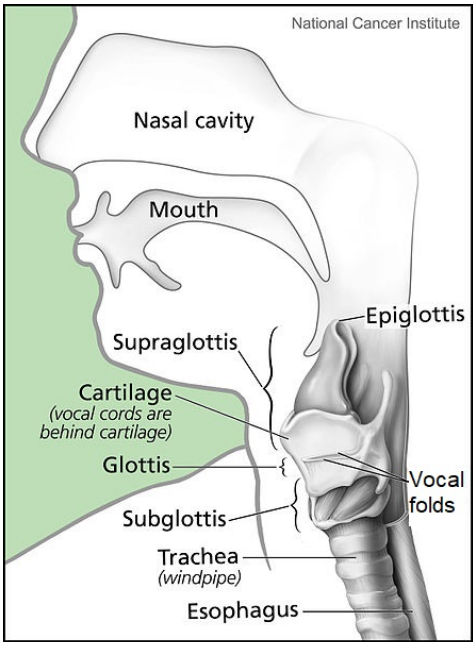


Figure 1.1: The main parts of the larynx (supraglottis, glottis, and subglottis) and other nearby structures, including the nasal cavity, mouth, cartilage, vocal folds, trachea and oesophagus.
(Hoofring 2003)

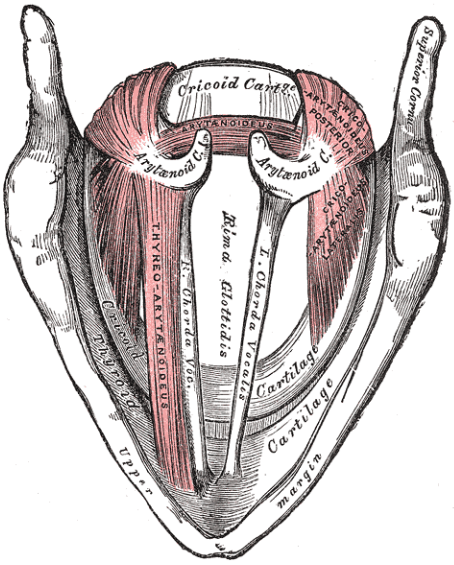


Figure 1.2: Muscles of larynx, seen from above. From Van Dyke (1918)

differ between cisgender male and female speakers, as well as across various cultures.

Figure 1.1 shows the main parts of the larynx and other nearby structures. Voice quality settings can be divided into supralaryngeal settings which involve the resonating cavities *above* the larynx, and laryngeal settings, or phonation, which relate to the configuration of the vocal folds (Laver 1980). Catford (1977) defines phonation as 'any laryngeal activity of speech that has neither initiatory nor articulatory function'. Laver (1980) identifies six basic phonatory settings: modal voice, falsetto, whisper, creak, harshness, and breathiness.

Figure 1.2 shows the muscles and cartilages of the larynx. Differences in phonation are produced by moving the muscles in the larynx, which changes the configuration of vocal folds (labelled on 1.2 as *chorda vocalis*) and causes them to vibrate in different ways. Modal voice is a neutral phonatory setting which other phonatory settings can be compared against, but which is rarely used in everyday speech. In modal voice, the arytenoid cartilages are brought together with moderate adductive tension and moderate medial compression, causing the vocal folds to vibrate periodically and the glottis to close completely with each vibration (Laver 1980). The auditory result of this is a voice quality with no audible friction or other auditory perturbations resulting from the larynx (Laver 1980).

Modal voice can be modified with ‘breathiness’ to produce breathy voice, where the vocal folds are still vibrating periodically, but do so less efficiently (Laver 1980: 132). Catford (1977: 99) likens the sound of breathy voice to sighing. In breathy voice, muscles in the larynx are held with low tension, resulting in the glottis being kept somewhat open along its length during vibration, and ultimately higher airflow through the glottis and audible friction (Laver 1980: 132). The auditory quality of breathy voice exists on a continuum with whispery voice, which also involves audible glottal friction, but a more substantial amount (Laver 1980: 133-134).

However, Laver (1980: 133) stresses that the two qualities are physiologically distinct: while breathy voice involves extremely weak medial compression in the vocal folds such that they do not meet along their length, whispery voice requires moderate to high medial compression which produces a triangular opening in the glottis adjacent to the arytenoid cartilages and higher audible friction. More recent work by Esling and Harris (2005) and Moisik et al. (2019) has investigated the differences between breathy and whispery voice further, finding that the main distinction between the two lies in the role of the epilarynx - that is, the ventricular folds and aryepiglottic folds in the upper portion of the larynx. In whispery voice, there is epilaryngeal constriction, which favours raising of the larynx. On the other hand, in breathy voice, there is no supraglottal constriction and a lack of overall tension, which favours lowering of the larynx.

However, as breathy and whispery voice are perceptually similar, and the present study is taking a perceptual approach to gender differences in voice quality, in this study I will group these two phonation types into one category, termed ‘breathy voice’, that can be defined as phonation that combines periodic voicing with a degree of glottal friction. This follows previous sociophonetic work on voice quality, such as Henton and Bladon (1985).

By contrast, in creak, the vocal folds are typically held tightly together so that they are thick and compressed, and the ventricular folds above them are also adducted (Laver 1980). Auditorily, it can be defined as a voice quality where separate low frequency glottal pulses can be identified auditorily (Henton and Bladon 1988: 4). According to Catford (1977: 98), creak can be said to sound like ‘a rapid series of taps, like a stick being run along a railing’. Varying laryngeal configurations can produce auditory effects that are commonly grouped under the term ‘creak’ (Keating et al. 2015), and various authors distinguish between these different categories. For example, Laver (1980) separates creak from ‘creaky voice’, a compound phonation type wherein creak is combined with modal voice. Here, I will follow Henton and Bladon (1988: 4) in grouping different types of creak into one category, as untrained listeners are unlikely to distinguish between different types of creak.

1.2.1 Measurement and analysis of phonation

Over the years, researchers have used a large number of different methods of measuring and analysing differences in voice quality.

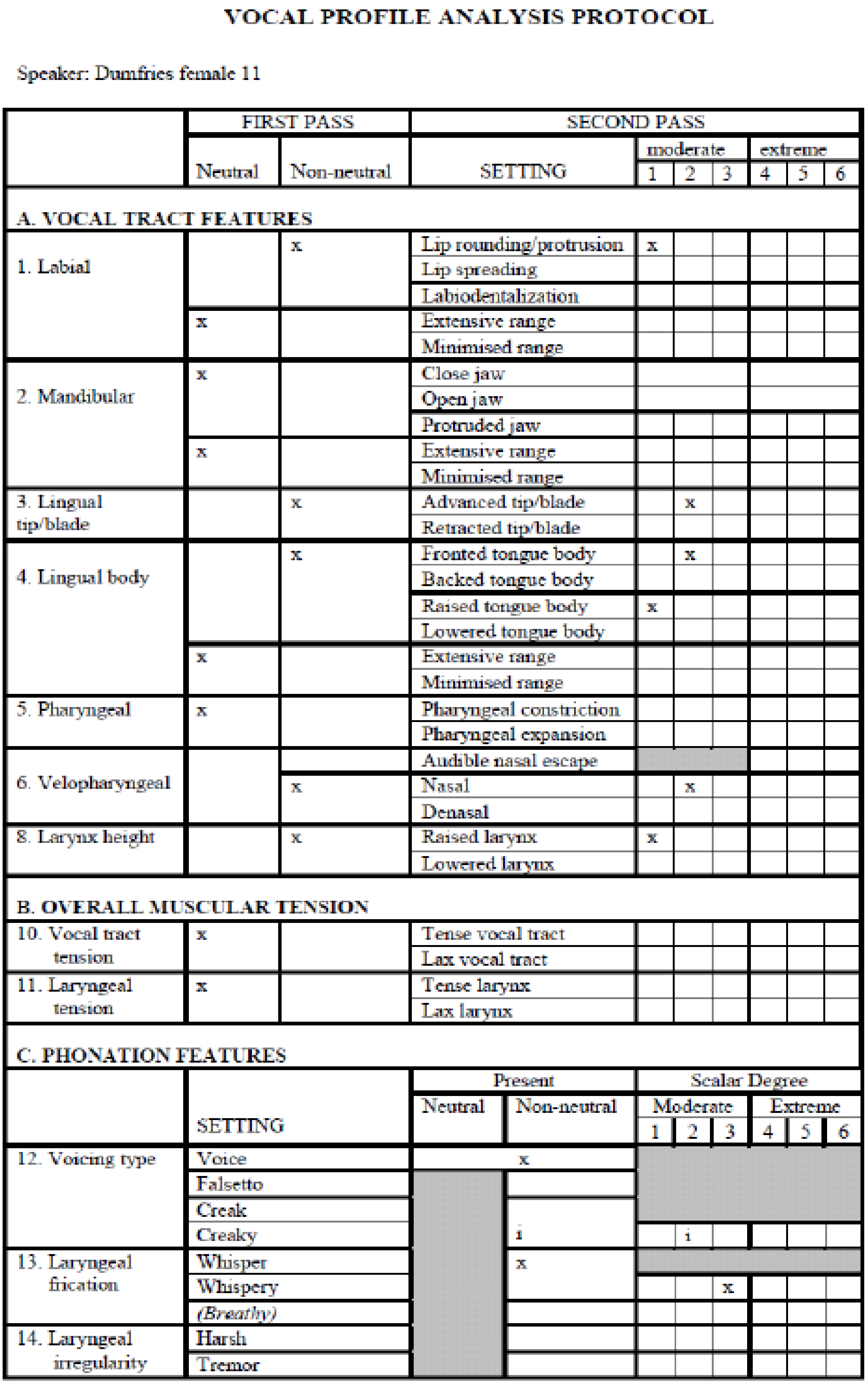
Auditory-perceptual approaches A common way of analysing voice quality is to use an auditory-perceptual method - to listen to speakers and follow some sort of protocol to make systematic judgements about speaker voice quality. While there are many auditory-perceptual approaches to analysing voice quality, one common method is the Vocal Profile Analysis (VPA) scheme, developed by Laver (1968, 1973, 1980) and described in detail in Beck (1988) and Laver (1980). The VPA considers voice quality in terms of the entire vocal tract; that is, not just phonation, but supra-laryngeal settings and overall settings of muscular tension. While it is an approach to the analysis of voice quality rather than an exact protocol (Beck 2005: 295), and variations exist on it, a typical implementation of it allows the judge to specify the voicing type (voice, falsetto, creak, creaky), degree of laryngeal frication (whisper, whispery, breathy), and laryngeal irregularity (harsh, tremor), as well as quantify the degree of each of these. Figure 1.3 gives an example of a completed Vocal Profile Analysis form for a speaker from Beck and Schaeffler (2015).

Studies may choose to take an auditory approach that focuses on a particular aspect of voice quality, rather than considering voice quality settings of the entire vocal tract. For example, Andrews and Schmidt (1997) asked listeners to rate how breathy each speaker was on a scale of 1-8, while Henton and Bladon (1988) and Yuasa (2010) focus specifically on the prevalence of creaky voice, and take the approach of counting the percentage of syllables in which creaky voice occurs for each speaker, coded auditorily and confirmed by looking at acoustic cues to creaky voice in the spectrogram.

Acoustic approaches Alternatively, it is possible to take a fully acoustic approach to the analysis of voice quality. Researchers have used a large number of acoustic measures in attempts to characterise differences in voice quality over the years; an overview of these measures can be found in Hillenbrand (2011), and Table 1.1 presents some of the most common measures.

Jitter and shimmer are two measures of perturbations of periodicity; jitter measures perturbation in fundamental period from cycle-to-cycle, while shimmer measures perturbation in the amplitude of adjacent pitch pulses. Jitter and shimmer are two of many measures that model voice quality in terms of the time domain, looking at the glottal pulses, their shape and differences between them.

It is also possible to investigate variation in the time domain by looking at the



VOCAL PROFILE ANALYSIS PROTOCOL

Speaker: Dumfries female 11

	FIRST PASS		SECOND PASS						
				moderate			extreme		
	Neutral	Non-neutral	SETTING	1	2	3	4	5	6
A. VOCAL TRACT FEATURES									
1. Labial		x	Lip rounding/protrusion	x					
			Lip spreading						
			Labiodentalization						
	x		Extensive range						
			Minimised range						
2. Mandibular	x		Close jaw						
			Open jaw						
			Protruded jaw						
	x		Extensive range						
			Minimised range						
3. Lingual tip/blade		x	Advanced tip/blade		x				
			Retracted tip/blade						
4. Lingual body		x	Fronted tongue body		x				
			Backed tongue body						
			Raised tongue body	x					
			Lowered tongue body						
	x		Extensive range						
			Minimised range						
5. Pharyngeal	x		Pharyngeal constriction						
			Pharyngeal expansion						
6. Velopharyngeal			Audible nasal escape	▨	▨	▨			
		x	Nasal		x				
			Denasal						
8. Larynx height		x	Raised larynx	x					
			Lowered larynx						
B. OVERALL MUSCULAR TENSION									
10. Vocal tract tension	x		Tense vocal tract						
			Lax vocal tract						
11. Laryngeal tension	x		Tense larynx						
			Lax larynx						

C. PHONATION FEATURES

		Present		Scalar Degree					
				Moderate			Extreme		
	SETTING	Neutral	Non-neutral	1	2	3	4	5	6
12. Voicing type	Voice	x		▨	▨	▨	▨	▨	▨
	Falsetto	▨		▨	▨	▨	▨	▨	▨
	Creak	▨		▨	▨	▨	▨	▨	▨
	Creaky	▨	i		i				
13. Laryngeal frication	Whisper	▨	x	▨	▨	▨	▨	▨	▨
	Whispery	▨				x			
	(Breathy)	▨							
14. Laryngeal irregularity	Harsh	▨							
	Tremor	▨							

Figure 1.3: An example of a completed Vocal Profile Analysis form, from Beck and Schaeffler (2015)

derivative of glottal flow, where the speech signal has been inverse filtered to remove the effect of the resonating cavities above the larynx, leaving only the source. For an overview of these measures, see Gobl and Ní Chasaide (2010). Time domain measures that look at the derivative of the glottal flow have the advantage of having a closer relationship to how differences are produced, while measures in the frequency domain are more closely linked to auditory perception (Gobl et al. 2019). Also known as spectral measures, these measures in the frequency domain model the voice source in terms of harmonics and the relationships between them, as well in terms of spectral noise.

One group of spectral measures looks at spectral tilt, which essentially measures ‘the degree to which intensity drops off as frequency increases’ Gordon and Ladefoged (2001: 397) by comparing the amplitude of different harmonics. Spectral tilt is related to the portion of the glottal cycle that the glottis is open (Holmberg et al. 1995) and how abruptly the vocal folds shut (Stevens 1977). As shown in Figure 1.5, harmonics drop off at different rates in different voice qualities, with the fall in energy at higher frequencies occurring least for creaky voice and most for breathy voice. Common ways of characterising this include the amplitude difference between the first harmonic (the fundamental, f0, or H1) and the second harmonic (H2), in a measure called H1-H2. In relation to modal voice, H1-H2 is higher for breathy voice and lower for creaky voice. Additional measurements of spectral tilt include H1-H3, H1-A1 (where A1 is the first formant) and H2-H4, which all follow similar patterns in terms of characterising the differences between modal, creaky and breathy voice. These measures are often written with asterisks, as H1*-H2*, to indicate that they have been corrected for formant values - this allows the measure to be used on different vowel qualities.

Another way to look at harmonic structure is to look at the spectrum of the log power spectrum - this is known as the cepstrum, a word which comes from rearranging the letters at the beginning of ‘spectrum’. Figure 1.4 shows the spectra of two vowels, one modal (a) and one moderately breathy (b), and their cepstrums in (c) and (d). Cepstral Peak Prominence (CPP) (Hillenbrand et al. 1994) is a measure of the amplitude of cepstral peak in the cepstrum which corresponds to harmonic regularity. More periodic signals have a more prominent cepstral peak and thus a higher CPP as in panel (c), reflecting a defined harmonic structure, while less periodic signals will have a less prominent cepstral peak as in panel (d). Higher CPP correlates strongly with listener ratings of breathiness.

Other measures look at how the proportion of harmonic, periodic energy in the signal relates to the amount to inharmonic noise. In modal speech, the signal is more periodic and so the harmonic energy is stronger than the inharmonic energy. The less modal speech becomes, the more inharmonic energy is found in the signal. Harmonics-to-Noise Ratio compares the ratio of harmonic to inharmonic spectral energy, so in

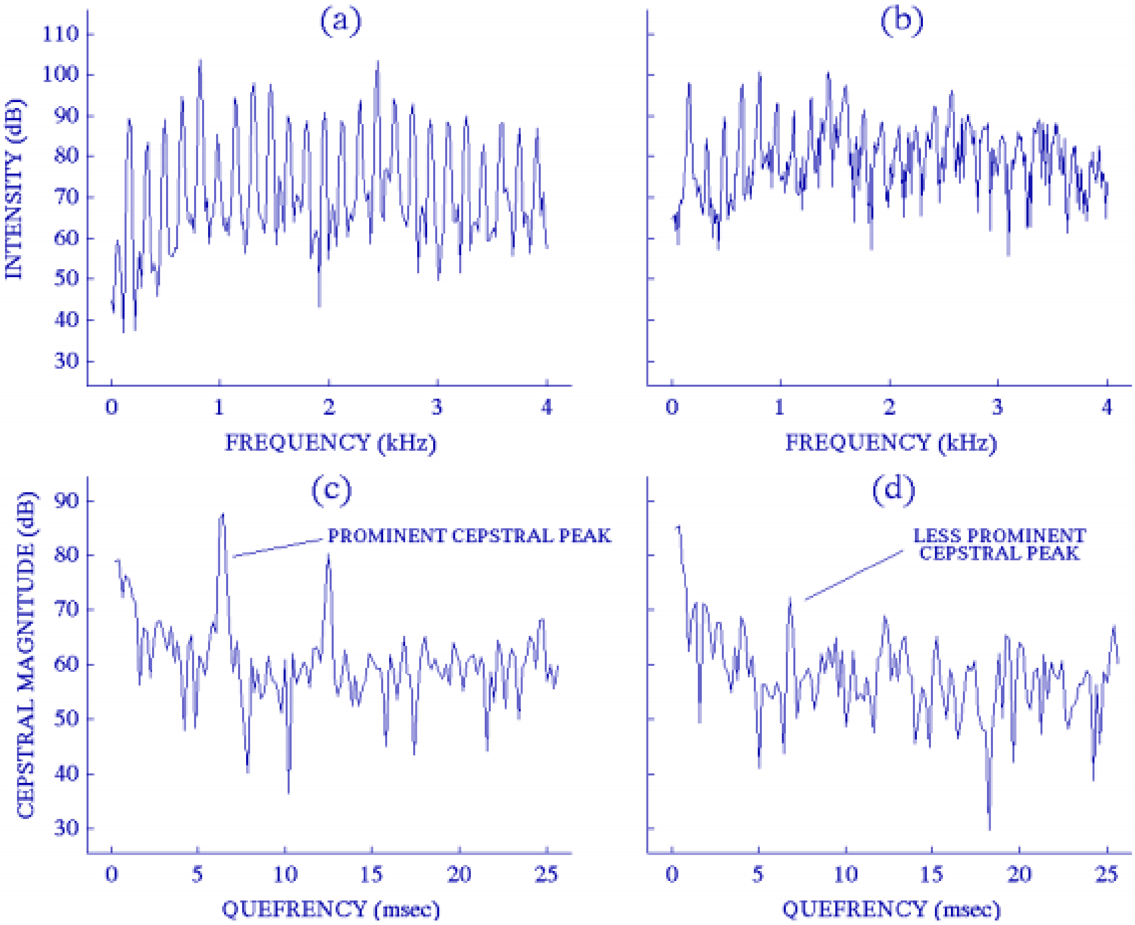


Figure 1.4: Panel (a) shows the spectra of a modal vowel. Panel (b) shows the spectra of a moderately breathy vowel. Panels (c) and (d) show the cepstrum of the vowels in (a) and (b) and the relative prominence of the cepstral peak. Taken from Hillenbrand (2011)

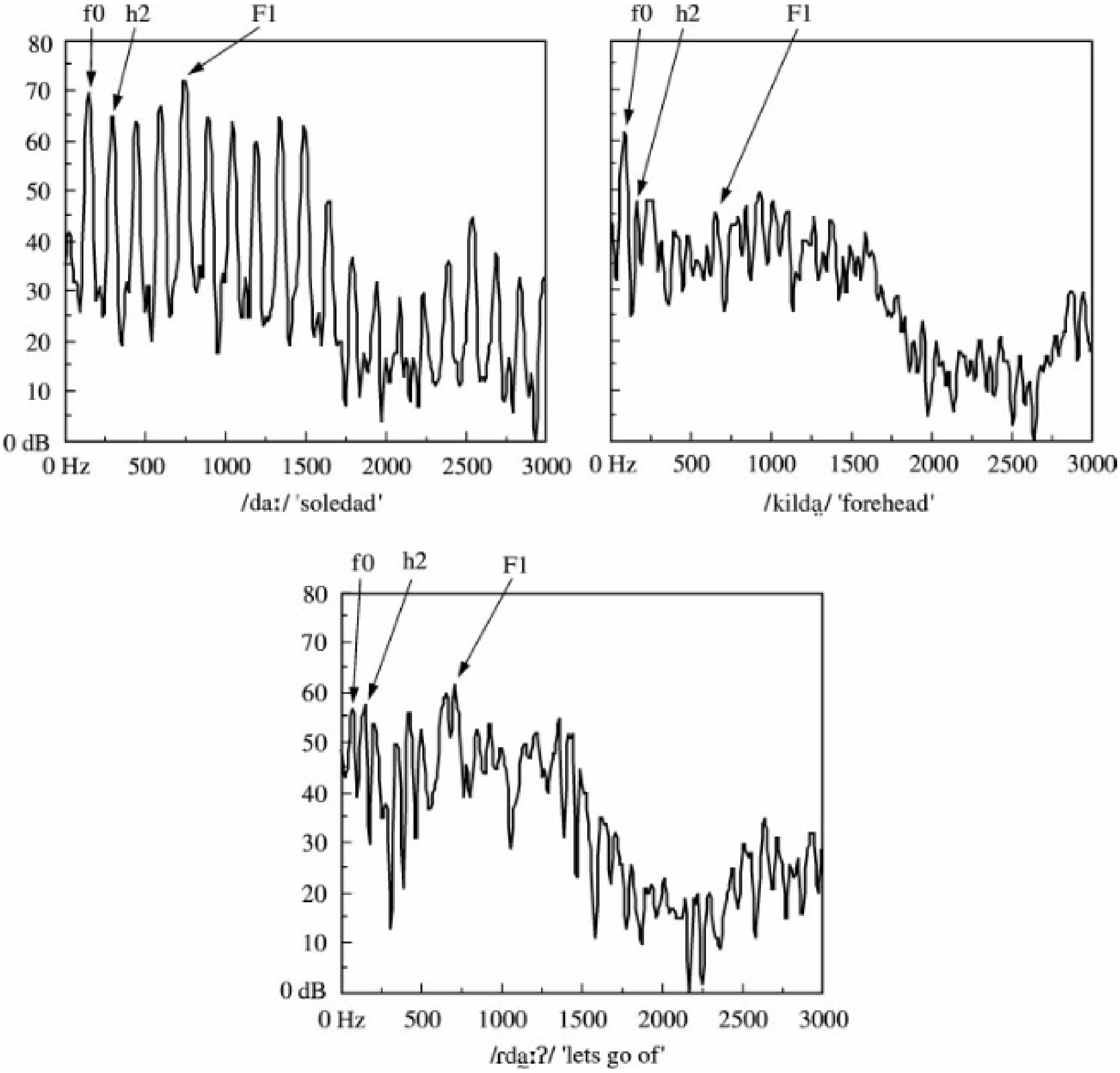


Figure 1.5: FFT spectra of modal, breathy, and creaky /a/ in the San Lucas Quiaviní Zapotec words /daː/ 'Soledad', /kilda̤/ 'forehead', and /ɾda̰ːʔ/ 'lets go of' (male speaker) from Gordon and Ladefoged (2001)

modal speech the HNR is higher in amodal speech HNR is lower. Specifically, HNR tends to be lower in creaky voice than in modal voice, due to the irregular f0 of creaky voice leading to an increase in noise, and even lower in breathy voice than in either modal or creaky voice, due to the presence of glottal friction in breathy voice.

There are numerous methods of calculating HNR, some which use the spectral domain (Yumoto et al. 1982), and some which use the cepstral domain (Krom 1993). Specific measurements of HNR, such as HNR05, HNR15 and HNR25, compare the ratio of inharmonic to harmonic energy at only certain frequencies. Noise-to-Harmonics Ratio (NHR) does the opposite of this, while other measures such as Voice Turbulence Index (VTI) and Soft Phonation Index (SPI) compare the ratio of harmonic to inharmonic energy at certain frequencies (*Multi-dimensional voice program model 5105* 2004).

An additional correlate of breathy voice is an increased F1 bandwidth. In breathy voice, incomplete glottal closure during a cycle leads to increased noise in the mid frequencies of the signal, and therefore a larger F1 bandwidth (Hansom 1997; Hansom and Chuang 1999).

As Kreiman et al. (2014) note, the number of measures developed to characterise voice quality has caused problems over the years. Some measures, such as jitter and shimmer, have been widely used, despite the fact that they do not correlate with listener's perception of the voice (Kreiman and Gerratt 2005).

Additionally, using only one type of acoustic measure can lead to problems interpreting results, because measures look at different aspects of the signal and tell the researcher about different things. Garellek (2019) demonstrates this problem in terms of H1-H2. He outlines how, while H1-H2 has been demonstrated to be useful for characterising voice quality differences, differences in H1-H2 are relative rather than absolute which can lead to difficulties in interpretation. For example, if we use H1-H2 in attempt to measure the difference in voice quality between two vowels, and vowel A has a higher H1-H2 than vowel B, it is difficult to tell based on this result alone if vowel A is modal and vowel B is creaky, if vowel A is breathy and vowel B is modal, or if vowel A is breathy and vowel B is creaky (Garellek 2019)

To resolve these kinds of issues, Kreiman et al. (2014) propose a psychoacoustic model of voice quality that integrates acoustic measurement of voice quality with perception and production, including measures of voice quality that they have found to be perceptually meaningful to listeners in their research. To establish the measures that were perceptually meaningful to listeners, they first investigated which measures were most successful in capturing variation in the shapes of glottal pulses and glottal source spectra (Kreiman et al. 2007), then investigated how perceptible each of these measures were to listeners (Kreiman et al. 2014).

Following from this, Garellek (2019) outlines that because voice quality differences can be described in terms of the degree of glottal constriction or spread, as well as the degree of noise, H1-H2, which correlates with glottal constriction, becomes more useful when combined with HNR, a measure of noise. Therefore, some research uses H1-H2 in tandem with HNR measures, combining a measure that captures relative differences between modal and non-modal speech (HNR) and a measure that captures relative differences between creaky, modal, and breathy speech (H1-H2), thus moving towards resolving the issue that Garellek (2019) identifies with using H1-H2 alone.

Fundamental frequency as a measure of voice quality F0 is often used in tandem with acoustic measures of voice quality, as differences in phonation are sometimes characterised or accompanied by differences in f0. This is especially the case with creaky voice, which is often characterised by its low and irregular f0. Dallaston and Docherty (2019) draws on research by Dorreen (2016) and propose a method for measuring the proportion of speech produced with creaky voice by using REAPER (Talkin 2015) to measure f0, then plotting all f0 values in a density plot. This generally results a bimodal distribution, where f0 values above the antimode consist of non-creaky voicing, and f0 values below the antimode consist of creaky voicing, allowing for calculation of the proportion of speech produced with creaky voice.

Instrumental approaches Alternatively, phonation can be investigated instrumentally. These include the use of airflow masks, which can be used to measure the flow of air through the vocal folds, as in Holmberg et al. (2010), and techniques like laryngoscopy, endoscopy, and stroboscopy, which involve imaging of the vocal folds by inserting a camera into the vocal tract, as in Palmer et al. (2012).

1.2.2 Gender differences in voice quality

In many languages, for example, Jalapa Mazatec and Gujarati, certain non-modal phonation types are used to express linguistic contrasts (Gordon and Ladefoged 2001). However, in English, evidence for phonation being used to express linguistic contrasts is only tentative (Di Paolo and Faber 1990), and instead differences in phonation can differentiate between individual speakers' voices, allow speakers to display stances in interaction, and signal information about the identity of the speaker (Podesva and Callier 2015). Information about speaker identity that can be signalled by phonation in varieties of English includes social class (e.g. Esling 1978; Stuart-Smith 1999) and ethnicity (e.g. Podesva 2013), as well as gender, and differences between speakers of these categories are often locally situated and intersect with one and other. For example, in Washington DC, Podesva (2013) found that African-American women use high rates of falsetto compared to African-American men and white American speakers. Meanwhile, in Edinburgh, Esling (1978) found that working class men produce more harsh voice than middle and upper class speakers. In terms of gender, the most widely documented differences in phonation exist in terms of breathy voice, modal voice and creaky voice, and therefore it is this aspect of phonation that will be investigated here. Table 1.1 presents a summary of previous research looking at gender differences in voice quality in varieties of UK and American English.

Gender differences in voice quality in the UK

Many studies investigating phonation in UK varieties of English have found that male speakers tend to be creakier, while female speakers tend to be breathier. Henton and

Table 1.1: Some commonly used acoustic measurements of voice quality

Parameter	**Description**	**Interpretation**
Jitter	Measure of cycle-to-cycle pertubation in fundamental period	Higher jitter suggests less modal, but no clear consensus
Shimmer	Measure of cycle-to-cycle fluctuation in the amplitude of adjacent pitch pulses	Higher shimmer suggests less modal, but no clear consensus
Harmonics-to-Noise Ratio (HNR)	Ratio of harmonic/periodic spectral energy to inharmonic spectral energy/noise	Originally developed to quantify hoarseness (Yumoto & Gould 1982); more generally, lower HNR suggests less modal
Noise-to-Harmonics Ratio (NHR)	Ratio of the inharmonic spectral energy/noise to harmonic/periodic spectral energy	Higher NHR suggests less modal
Voice Turbulence Index (VTI)	Ratio of spectral inharmonic high-frequency energy (2800-5800 Hz) to the spectral harmonic energy in the range 70-4500 Hz	Higher VTI suggests less modal
Soft Phonation Index (SPI)	Ratio of the low-frequency harmonic energy (70-1600 Hz) to high-frequency energy (1600-4500 Hz)	Higher SPI correlates with incomplete vocal fold closure and thus, breathiness
Cepstral Peak Prominence (CPP)	Amplitude-normalized measure of the amplitude of the cepstral peak corresponding to the harmonic regularity	Less periodic signals have a lower CPP; Has been shown to correlate with breathiness
Bandwidth of the first formant (B1)	Bandwidth of the first formant	Increased B1 in breathy voice due to incomplete glottal closure and increased noise
H1-H2	Amplitude difference of the first to the second harmonic	Higher values in breathy voice, lower values in creaky voice
H1-H3	Amplitude difference of the first to the third harmonic	Higher values in breathy voice, lower values in creaky voice
H1-A1	Amplitude difference of the first harmonic to the first formant	Higher values in breathy voice, lower values in creaky voice
H2-H4	Amplitude difference of the second to the fourth harmonic	Higher values in breathy voice, lower values in creaky voice

Bladon (1985) investigated breathy voice in the speech of Northern English speakers who have since moved to the south of England and modified aspects of their speech (Modified Northern; MN) and speakers of Received Pronunciation (RP). They found that among both MN and RP speakers, H1-H2 was significantly higher in female speakers than in male speakers, indicating that female speakers were more breathy. Henton and Bladon (1988) also investigated creakiness in speakers of MN and RP, coding for creakiness auditorily. They found that male speakers were significantly creakier than female speakers. Additionally, they found differences within genders and between the two groups of speakers: male and female speakers of MN were significantly creakier than their same-gender RP counterparts.

These results are also supported by Stuart-Smith (1999)'s sociophonetic study of voice quality in Glasgow, which used the Vocal Profile Analysis (VPA). As in Henton and Bladon (1988), male speakers were found to be significantly creakier than female speakers. Female speakers were found to be significantly more whispery than male speakers. However, there was no difference in the use of breathy voice. Stuart-Smith (1999) notes that, 'If 'whispery voice' is to be equated with Henton and Bladon's 'breathiness' [22], this again would seem to agree with their observations'. However, the fact that Stuart-Smith (1999) identifies a gender difference in whispery voice but not in breathy voice suggests that the cultural meanings of these two voice qualities are not equivalent.

Beck and Schaeffler (2015) also used the VPA in their investigation of voice quality among among 76 Scottish adolescents from different parts of Scotland. While they did not find increased whispery voice or breathiness among female speakers, they did find that male speakers were significantly creakier.

In a study that looked at phonation in a sample of over 1000 speakers from across the UK, Gittelson et al. (2018) also found evidence that female speakers are breathier while male speakers are creakier. Combining H1*-H2* (a formant-corrected measure of H1-H2) with a measure of Harmonics-to-Noise Ratio (HNR15), they find that male speakers tend to have a lower H1*-H2*, indicating that they are less breathy and more creaky, as well as that they have a lower HNR15, indicating that they are less modal. They used data collected through the English Dialects App from speakers from across the UK and do not find any systematic geographic variation in their results; however, they did not check for an interaction of gender and dialect so it is possible that this tendency may vary between different UK varieties of English.

Szakay and Torgersen (2015) investigated differences in phonation using H1-H2 among speakers living in Hackney in inner city London and Havering in outer London, and speakers with local family roots (Anglo) compared with those that were children or grandchildren of immigrants (non-Anglo). Among Anglo speakers from Havering, male speakers had a lower H1-H2 while female speakers had a higher H1-H2, indicating that male speakers are creakier while female speakers are breathier. However, in Hackney, male Anglo speakers had a higher H1-H2, while female Anglo speakers had the lowest values for H1-H2 out of all the groups. Furthermore, both male and female non-Anglo speakers from Hackney exhibited high values of H1-H2 and male speakers having a higher H1-H2 than female speakers, indicating breathy voice.

However, Szakay and Torgersen (2019) revisit this data again using a REAPER analysis of creaky voice. While they again find differences between their groups in terms of creaky voice use, the groups that are indicated to be the most creaky when analysed using H1-H2 do not match those measured by REAPER analysis. In Hackney, Anglo female speakers use the least amount of creaky, followed by Anglo male speakers, non-Anglo female speakers, and finally non-Anglo male speakers. Anglo speakers in Havering were more creaky overall than any groups in Hackney; and the trend for male speakers to be creakier was reversed, with Anglo female speakers in Havering being creakier than Anglo male speakers in Havering.

Taken together, the findings of Szakay and Torgersen (2015, 2019) not only stress the fact that gender differences in voice quality are culturally-specific, but also suggest that a breathier voice, indicated by a higher H1-H2, does not necessarily mean less creaky voice. They also reinforce the importance of considering measures of H1-H2 in tandem with other measures of voice quality.

Gender differences in voice quality in North America

Syrdal (1996) looked at the characteristics of speech from telephone conversations in speakers of American English, including voice quality through H2-H1 (a measure of spectral tilt which takes the difference between the amplitude of the second harmonic and the first, rather than the other way around as in H1-H2) and auditory coding of creaky voice. While they did find that female speakers had a lower H2-H1, indicating they were more breathy, they were surprised to find that unlike what had been found in previous research, female speakers in fact used more creak than male speakers did.

Yuasa (2010) investigated creaky voice usage in female speakers of American English in California, comparing their use to female speakers of Japanese and male speakers of American English. Through auditory coding of creaky voice, she found that female speakers used more creaky voice than either Japanese-speaking women or American English-speaking men.

Similarly, looking at the use of voice quality among African American and white residents of Washington DC, Podesva (2013) finds that women use more creaky voice than men. Contrary to most research that looks at the use of breathy voice, Podesva

(2013) also finds that there is no difference in the use of breathy voice use between male and female speakers, among neither African American or white participants. However, defining whispery voice as voiceless, he did find that white men used low amounts of whispery voice in comparison to white female speakers and both male and female African American speakers.

Wolk et al. (2012) and Abdelli-Beruh et al. (2014) investigated the use of creaky voice among 34 male and 34 female speakers of American English, and find that creaky voice was common among female speakers but rare among male speakers. Furthermore, Becker et al. (2017) investigated creaky voice in American English usage among a diverse gender sample, including not only cisgender male and female participants but also transgender men, transgender women, and non-binary speakers. They find that creaky voice was used more by cisgender female speakers than cisgender male speakers, it did not pattern according to gender identity overall when transgender and non-binary speakers were included.

Comparing gender differences in voice quality between American and UK varieties of English show that gendered meanings of creaky voice are culturally specific. Given that in some languages voice quality has functions such as expressing linguistic contrasts (Gordon and Ladefoged 2001), if we were to look beyond English, it is likely that this would become even more apparent.

Explanations for gender-based voice quality differences

There is some evidence that the increased prevalence of breathy voice in female speakers could be a result of sexual dimorphism in the vocal tract. Female vocal folds are, on average, around 20-30% thinner vertically than male vocal folds (Hollien 1960). In Titze (1989)'s computational model of vocal fold contact in male and female speakers, this difference in vocal fold thickness results in incomplete vocal fold contact towards the arytenoid cartilages during vibration, and thus produces audible friction. However, as we have seen, female speakers are not breathier than male speakers across all varieties of English (Podesva 2013; Szakay and Torgersen 2015), suggesting that if there is a physiological influence, it is one that speakers can overcome.

According to Simpson (2009), there do not appear to be any anatomical or physiological reasons for male speakers to be creakier. Beck and Schaeffler (2015) have speculated that the laryngeal instability caused by hormonal changes during puberty may be a reason for increased creakiness in adolescent male speakers; however, this would not explain why in American English, female speakers have been found to be particularly creaky. Again, this indicates that speakers can compensate for the potential influence of vocal tract physiology.

However, while there is not a clear physiological reason for male speakers to be creakier, Podesva and Kajino (2014) suggest that there may be an iconic link between creaky voice and masculinity, regardless of the gender of the speaker using it. Podesva and Kajino (2014) propose that the low f0 of creaky voice is heard as resembling the low f0 that male speakers tend to use, and is thus interpreted as masculine. However, they stress that this meaning of creaky voice is mediated by ideology, allowing it to become associated with femininity through use in women's speech, as has occurred in the US.

Furthermore, the large amount of variation between gender and sex categories, as well as across different cultures and language varieties, shows that, as Podesva and Callier (2015: 177) note, while physical differences in the vocal tract may influence phonation, these differences 'merely define the envelope of vocal possibilities rather than determine where within the possible range women and men will fall'.

Table 1.2: Table summarising previous research looking at gender differences in voice quality production

Article	**Study purpose**	**Method of measuring VQ**	**Speaker demographics**	**Results/conclusions**
Becker, Khan and Zimman 2017	Explore creaky voice use within a diverse gender sample to investigate the social meanings of creaky voice	Auditory coding	43 American English speakers stratified across current gender identity (women, men, non-binary), sex assignment at birth (female, male), and exposure to testosterone	Cisgender women use more creaky voice than cisgender men, but overall there is no pattern according to sex or gender
Henton & Bladon 1985	Investigate use of breathy voice in non-pathological speech in British English and how this patterns according to gender and dialect	H1-H2	26 Received Pronunciation speakers (20 female, 16 male); 25 Northern English speakers (12 female, 13 male)	H1-H2 higher in female speakers, indicating that women are breathier than men
Henton & Bladon 1988	Investigate use of creak in non-pathological speech in British English and how this patterns according to gender and dialect	Auditory coding	40 Received Pronunciation speakers (20 female, 20 male); 40 Modified Northern speakers (20 female, 20 male)	Men use more creak than women
Stuart-Smith 1999	Describe voice quality in Glaswegian English and investigate whether there is any evidence for the stereotypical Glaswegian voice quality	Vocal Profile Analysis	32 speakers (male and female, adults and children, working class and middle class) from Glasgow	Women use more whispery voice; Men use more creaky voice

Table 1.2: Table summarising previous research looking at gender differences in voice quality production

Article	Study purpose	Method of measuring VQ	Speaker demographics	Results/conclusions
Beck & Schaeffler 2015	Identify voice quality markers of gender and geographical background among Scottish adolescents	Vocal Profile Analysis	76 adolescent speakers (31 male; 45 female) from different regions of Scotland	Male speakers were creakier
Gittelson, Li & Leeman 2018	Investigate how demographic factors such as gender, age, and place of origin affect phonation	H1*-H2* and HNR	1000+ speakers from across the UK	Male speakers showed a lower H1*-H2* and lower HNR than female speakers; suggests male speakers creakier and female speakers breathier
Podesva 2013	Investigate voice quality among residents of Washington DC	Auditory coding	32 speakers (half white, half African American, and with each group, half male, half female)	White male speakers used less whispery voice than any other group; female speakers were creaky than male speakers
Szakay & Torgersen 2015	Investigate voice quality in London English using acoustic measures and compare VQ between inner London, outer London, anglo and non-anglo speakers	H1*-H2* and f0	42 speakers (28 from Inner London, '18 male and 9 female'; 14 from outer London, 7 male and 7 female)	Anglo Havering: male speakers more creaky, female speakers more breathy; Anglo Hackney speakers: male speakers more breathy, female speakers more creaky; non-Anglo Hackney: male and female speakers found to be breathy (but male speakers esp. breathy)

Table 1.2: Table summarising previous research looking at gender differences in voice quality production

Article	**Study purpose**	**Method of measuring VQ**	**Speaker demographics**	**Results/conclusions**
Szakay & Torgersen 2019	Reanalyse data from Szakay & Torgersen (2015) with REAPER to investigate creak and f0 in this population	REAPER f0 analysis	Same as in Szakay & Torgersen (2015)	From least creak usage to most creak usage: Hackney Anglo Female, Hackney Anglo Male, Hackney non-Anglo female, Hackney non-Anglo male, Havering Anglo Male, Havering Anglo Female
Syrdal 1996	Describe acoustic characteristics of spontaneous conversational telephone speech	H2-H1 and auditory coding of creak	160 speakers of American English (80 male, 80 female)	Higher H2-H1 for male speakers than for female speakers, suggesting male speakers more laryngealised, but conflictingly, more auditorily creaky tokens for female speakers
Yuasa 2010	Examine usage of creaky voice in conversational speech in American and Japanese speakers	Auditory coding confirmed acoustically by referring to spectrogram	23 speakers (11 male, 12 female) of American English (California); 10 female speakers of Standard Japanese	American women use more creaky voice than American men and Japanese women
Wolk, Abdelli-Beruh & Slavin 2012	Examine the use of vocal fry in young adult speakers of American English	Auditory judgments; f0 measures, jitter local, shimmer local, and HNR	34 female speakers of Standard American English	Vocal fry was commonly used by female speakers in sentence-reading, but not in isolated vowels

Table 1.2: Table summarising previous research looking at gender differences in voice quality production

Article	**Study purpose**	**Method of measuring VQ**	**Speaker demographics**	**Results/conclusions**
Abdelli-Beruh, Wolk & Slavin 2014	Examine the use of vocal fry in young adult speakers of American English	Auditory judgments	34 male speakers of Standard American English	Vocal fry was rarely used by male speakers in either sentence-reading or isolated vowel production

1.3 Phonation and gender perception

1.3.1 What we know

Despite the fact that findings of differences in voice quality between male and female speakers are culturally specific, voice therapy protocols for voice masculinization and feminization therapy for English-speaking transgender and gender non-conforming people often include changes to voice quality as potential goals in therapy. For example, the Royal College of Speech and Language Therapists (2018) competency framework for speech and language therapists conducting trans and gender-diverse voice and communication therapy lists being able to deliver interventions in voice quality for voice masculinization and feminization as one of the clinical skills that therapists should develop when working with trans clients.

Furthermore, Davies and Goldberg (2006) includes the adoption of breathy voice as a goal in voice feminization, saying that not only is breathy voice perceived as more feminine in itself, but that a breathy voice quality will also help to decrease loudness, if this is desired, and to 'soften' hard vowel and consonant onsets which are often considered masculine, although what is meant by hard vowel and consonant onsets is never fully explained. Sawyer (2019) also notes that breathy voice has often been a target in feminizing voice therapy, but suggests that it may not be an efficient method of feminising the voice because breathy voice may decrease loudness, and thus might increase vocal fatigue. Davies et al. (2015) also notes that breathy voice is a common goal in voice therapy protocols for feminizing voice therapy, but stresses that more research is needed to confirm the effect of voice quality on the perception of transgender women's voices. Establishing to what extent voice quality contributes to gender perception thus has important implications for voice therapy for transgender individuals, as self and listener ratings of gender in the voice have been found to be correlated with transgender individuals' quality of life (Hancock et al. 2011).

Often cited as supporting the finding that breathy voice contributes to the perception of femininity is Van Borsel et al. (2009)'s study on Dutch speakers and listeners. They investigated whether breathy voice increased the perception of femininity by looking at the perception of breathy and non-breathy productions of /a/ in cisgender women. They found that breathy productions were judged as more feminine than non-breathy productions, but stressed that further work was needed to establish the degree of breathy voice needed to increase the perception of femininity and investigate whether this would also apply to the speech of transgender women.

However, as the present study looks at the contribution of voice quality to gender perception among English speakers, I will mostly focus here on previous work conducted on varieties of English, or, in the absence of the authors specifying the language or variety spoken by participants, work that can reasonably be assumed to have been conducted on a variety of English given the location of the university that researchers were based at or which they mention recruiting participants from.

Several studies looking at varieties of English give evidence to suggest that voice quality contributes to gender perception. Addington (1968) investigated the relationship between seven different voice qualities produced by two male and two female speakers, and a range of listener judgements, including masculinity and femininity, in what we can only assume are American English speakers and listeners. A breathy voice in women resulted in voices being rated as more feminine; however, a breathy voice in men did not result in a change in how masculine or feminine a voice was rated.

Looking at the voices of 11 self-described heterosexual intermittent cross-dressers, Andrews and Schmidt (1997) also found that breathy voice increases the perception of femininity in the voice, presumably in American English. They compared participants' voices when speaking in their male mode to their voices when speaking in a female mode. Seven out of the 11 speakers were perceived as significantly breathier when speaking in their female voice than in their male voice. Additionally, increased breathy voice was one of three characteristics that was most associated with increased perception of the voices as feminine when speaking in their female voice.

These findings are somewhat supported by studies that investigate vocal fold function in transgender women instrumentally. Presumably looking at American English, Palmer et al. (2012) used endoscopic and stroboscopic procedures to investigate vocal fold configurations in transgender women attempting to achieve a female-sounding voice. They found that most participants exhibited incomplete glottal closure and a posterior glottal gap, suggesting that they used a breathy voice. Furthermore, participants who exhibited this configuration were more likely to report passing as female in their everyday lives and some of their voices successfully elicited 'female' responses when their voices were presented in a perceptual experiment. Similarly, Gorham-Rowan and Morris (2006) used inverse filtering of airflow signal to investigate vocal fold function in transgender women, presumably speakers of American English. They found that when producing /a/ in their female voices, transgender women exhibited a significant increase in maximum airflow compared to when they produced /a/ in their male voices. They conclude that this strategy is effective for some participants, who were rated as more feminine in their female voices than in their male voices, but not for all participants.

Looking speakers of Australian English, Holmberg et al. (2010) also found that transgender women produced high rates of airflow when when this was measured using airflow masks. However, they found that the voices of transgender women had low rates of breathy voice, and that the amount of breathy voice had no relationship to whether their voices were rated as more male or more female. On the other hand, they found that transgender women who reported themselves having a 'croaky' voice were rated as less female and reported lower satisfaction with their voice, leading the authors to suggest that avoidance of creaky voice may help transgender women to develop a female-sounding voice.

Other studies have investigated the relationship between differences in acoustic measures of voice quality and the perceived gender of a voice, and found mixed results. Looking at Mandarin and American English speakers, Bishop and Keating (2012) set out to investigate whether voice quality conveyed information about whether a sound was high or low in a speaker's overall pitch range. They found that voice quality only conveyed this information indirectly, i.e. through giving listeners a cue to speaker gender. As well as finding that listeners used f0 and vowel formant frequencies to make decisions about speaker gender, they found that voice quality cues also contributed. Specifically, they found that higher H1*-A3*, an acoustic correlate of breathy voice, increased the likelihood that speakers would be heard as female. They also found that at a lower f0, a higher H2*-H4* increased the likelihood that speakers would be heard as male; they suggest that this unexpected finding might reflect the absence of a creaky voice quality that female speakers would be expected to produce at a lower f0.

Looking at North American English, Porter (2012) investigated the relationship between a range of voice quality measures and perception of speaker gender in the speech of transgender women, cisgender men, and cisgender women. While she found that a range of voice quality measures correlated with gender perception, the importance of different voice quality parameters differed between carrier phrases and sustained vowels. In sustained vowels, several parameters associated with breathy voice were correlated with increased perceptions of femininity. These were voice turbulence index, H1*-H3* and soft phonation index. However, this was not the case in carrier phrases. Instead, H1*-H2* and cepstral peak prominence, different parameters associated with breathy voice, were correlated with perceptions of increased femininity.

Skuk and Schweinberger (2014) morphed different acoustic cues to gender on a continuum from male to female to establish their relative contribution to listener judgements of speaker gender using the software TANDEM-STRAIGHT (Kawahara and Morise 2011). The resulting stimuli differed in terms of f0 and vowel formant frequencies along with two voice quality parameters: spectrum level, which reflected formant bandwidths and global spectral tilt, and aperiodicity. They found that when spectrum level was morphed on a continuum from male to female, it had a significant effect on the perception of speaker gender, firstly in an experiment where f0 and vowel formant frequencies were held constant at an ambiguous value and then in a second experiment where they were co-varied alongside it. However, they also found that aperiodicity had no effect, and that the contribution of spectrum level was less than that of f0 and formant frequencies. However, Skuk and Schweinberger (2014) looked at German, rather than English, and it may be that the contribution of voice quality to gender perception differs between languages and varieties.

On the other hand, looking presumably at American English, Owen and Hancock (2010) found no relationship between acoustic measures of breathy voice and either self or listener rated vocal femininity in the speech of transgender women. However, they investigated different measures: noise-to-harmonics ratio, voice turbulence index, shimmer and jitter. They also noted that this finding did not necessarily indicate that breathy voice and other aspects of voice quality were not related to listener perception of femininity; rather, they suggest that this may be related to limitations of the measures they used, and additionally may stem from the fact that the voice quality of their participants is relatively homogeneous, as all the transgender women in their study had completed feminizing voice therapy.

Looking presumably at American English, King et al. (2012) also looked at jitter and shimmer when examining the relationship between gender perceptions and a range of acoustic voice measures in transgender and cisgender women, and found no evidence that theses measures had a relationship with gender perceptions. Additionally, in a study examining whether clear speech contributed to gender perception in American English, Booz and Ferguson (2016) found no relationship between cepstral peak prominence and the perception of speech as more masculine or feminine.

Several studies have investigated the perception of creaky voice in American English speakers, with a view to establish why female speakers are using more creaky voice than male speakers. As well as looking at production, Yuasa (2008) investigated listener perception of creaky voice in American English. While she did not directly ask listeners whether they thought it sounded ‘masculine’ or ‘feminine’, she suggests that rather than being associated with masculinity and authority as was once thought, creaky voice is emerging as a new feminine voice quality that indexes being educated, urban-oriented and upwardly mobile. The potential for creaky voice to index femininity in American English is to an extent supported by the findings of Zimman (2013), who investigated the characteristics of gay-sounding speech in American English and found a correlation between creaky voice and listener perception of a speaker as gay-sounding.

Greer (2015) and Lee (2016) both investigated the perception of creaky voice in American English speakers. Contrary to the speculation of Yuasa (2010) that creaky voice was emerging as a feminine voice quality, they both find that creaky voices were rated as more masculine by listeners than non-creaky voices. Greer (2015) additionally found that breathier voices in women were perceived as more feminine compared with less breather or non-breathy voice qualities.

While most studies investigating the contribution of voice quality to gender perception have not taken the impact of listeners' sociolinguistic background into account, Foulkes et al. (2011) investigated how listeners judge gender in children's speech and compared between listeners from the UK and the US. They found that voice quality had no significant effect on UK listeners' perception of gender in children, but that in certain linguistic contexts, breathy voice led listeners from the US to be more likely to categorise the speaker as a girl.

The aims, methods, and results of these studies are summarised in Table 1.3.

1.3.2 What we don't know

Overall, while there is some evidence that voice quality contributes to the perception of gender in transgender women, this evidence is not conclusive. Reviewing existing literature on the contribution of voice quality to gender perception, both Davies et al. (2015) and Leung et al. (2018) argue that more work is needed to establish the role of voice quality to the perception of gender, and that there is insufficient evidence that breathy voice contributes to increased perception of a voice as feminine or female. Several factors may have contributed to the fact that previous research has found inconsistent and inconclusive results. The first, highlighted by Leung et al. (2018), is that studies use different methods of quantifying voice quality, and both acoustic and perceptual approaches to this have limitations. Leung et al. (2018) outline how studies which use a perceptual-auditory method of quantifying voice quality have tended to find a relationship between voice quality and perceived gender, while studies that have used acoustic measurements of voice quality have tended to find that there is no relationship. Further, most studies that have investigated this issue have looked at correlations between perceptual ratings or acoustic measures of voice quality and listener ratings of speaker gender, rather than taking a more experimental approach and attempting to manipulate voice quality. For this reason, it is difficult to separate the contribution of voice quality from that of other aspects of speech that might co-occur with it. By synthesising voice quality differences, the present study seeks to resolve some of the difficulties that arise in correlational studies using either auditory-perceptual or acoustic measures of voice quality, by increasing the amount of control over the stimuli.

A further potential reason for conflicting results in previous studies may be sociolinguistic differences between speakers and listeners of different dialects. As we have seen, voice quality differences between cisgender male and female speakers differ between speakers of different varieties of English. For example, in London, Szakay and Torgersen (2015, 2019) found that gender-based voice quality differences varied between people from different parts of London and with different ethnic backgrounds. It follows that listeners' sociolinguistic background may also affect how they interpret voice quality differences in voices; indeed, this was the case in Foulkes et al. (2011)'s study on listener perception of gender in children's voices. In the present study, I will address this issue by controlling for listener background in two experiments where listeners will hear stimuli recorded by a speaker of Scottish English, in which male speakers have been found to be creakier, while female speakers have been found to be more whispery. In the first, all participants will be speakers of Scottish English originating from the central belt of Scotland. In the second, North American participants will be recruited to examine how differences in listener background influence voice quality perception.

A further reason for conflicting results in previous studies may be the fact that the length of stimuli used differs between different studies. While many studies investigate isolated vowels (Bishop and Keating 2012; Gorham-Rowan and Morris 2006; Palmer et al. 2012; Van Borsel et al. 2009), others investigated isolated syllables (Holmberg et al. 2010; Skuk and Schweinberger 2014), and some looked at longer portions (Addington 1968; Andrews and Schmidt 1997; King et al. 2012). Porter (2012) compared the contribution of voice quality to gender identification across different types of stimuli and found important differences between isolated vowels and longer types of stimuli. While the present study will not compare the contribution of voice quality to gender identification between different types of stimuli, by using short sentences as the stimuli this should help to increase the authenticity of the stimuli and give a better idea of the contribution of voice quality to gender identification in everyday speech than using isolated syllables or vowels.

Additionally, as noted by Leung et al. (2018), studies have quantified 'perceived gender' on a variety of different rating scales. These include binary, categorical, and continuous scales of 'masculine'-'feminine', 'male'-'female', or 'man'-'woman'. Listeners may interpret the meanings of each of these terms differently, and thus studies which use different scales may produce conflicting results. The present study seeks to establish whether the type of response elicited affects whether voice quality contributes to gender perception, and will compare how voice quality affects listener responses of 'man', 'woman' or 'neither', compared to how it affects listener responses of 'not at all masculine' - 'masculine' and 'not at all feminine' - 'feminine' on a slider with values between 1-100.

Table 1.3: Table summarising previous research looking at the contribution of voice quality to gender perception

Article	**Study purpose**	**Method of producing/measuring VQ**	**Stimuli characteristics**	**Speaker demographics**	**Speaker gender rating scale**	**Listener demographics**	**Results/conclusions**
Addington 1968	Examine the relationship of vocal characteristics to perceived personality traits	Speakers instructed to produce different voice qualities	Read passage	Two male and two female speakers - language and variety unspecified, but likely to be American English	7-point scale from 'masculine' to 'feminine'	Approximately twenty listeners of unspecified demographics	Among female, but not male, speakers, breathy voice perceived as more feminine
Andrews & Schmidt 1997	Examine differences in perceptual and acoustic characteristics of the voices of speakers presenting in masculine and feminine modes	Speakers asked to read a passage in a masculine voice and in a feminine voice; listeners rated voices on an 8-point scale from breathy-full	Read passage	11 self-described heterosexual cross-dressers; language and variety unspecfied, but likely to be American English	8-point scale from masculine-feminine	88 students majoring in speech & hearing sciences of unspecific demographics	7 out of 11 speakers rated more breathy in their feminine voice than in their masculine voice; increased breathy voice in feminine voice increased perception of voice as feminine
Bishop & Keating 2012	Examine what information listeners use to tell if a sound is high or low in a speaker's overall range	Voice quality measures: CPP, H1*-H2*, H1*-H3*, H1*-A1*, H2*-H4*	Isolated vowels	Ten adult native speakers of English and Mandarin (5 male and 5 female for each language)	Binary choice of 'male' or 'female'	23 native speakers of English and 23 native speakers of Mandarin	Higher H1*-H3* (increased breathy voice) increased likelihood of speakers being heard as female; higher H2*-H4* at low f0 increased likelihood that speakers would be heard as male

Table 1.3: Table summarising previous research looking at the contribution of voice quality to gender perception

Article	**Study purpose**	**Method of producing/measuring VQ**	**Stimuli characteristics**	**Speaker demographics**	**Speaker gender rating scale**	**Listener demographics**	**Results/conclusions**
Booz & Ferguson 2016	Examine the relationship between clear speech style and vocal femininity	Cepstral peak prominence smoothed	Carrier phrases	41 talkers from the Ferguson Clear Speech Database	100 point scale from masculine to feminine	17 listeners. 8 female, 9 male.	No relationship between CPPS and perceived femininity
Gorham-Rowan & Morris 2006	Describe glottal airflow characteristics of the voices of transgender women and determine whether aerodynamic measures correlate with listener perception of femininity in transgender women	Inverse filtering of airflow signal to estimate vocal fold function	Isolated vowels	13 transgender women aged 24-55; language and variety unspecified, but likely to be American English	100-mm line from masculine to feminine	14 speech-language pathology students	Transgender women increased maximum airflow when speaking in their female voices, suggesting more breathy; this had an effect on listener ratings of femininity for some, but not all, participants

Table 1.3: Table summarising previous research looking at the contribution of voice quality to gender perception

Article	Study purpose	Method of producing/measuring VQ	Stimuli characteristics	Speaker demographics	Speaker gender rating scale	Listener demographics	Results/conclusions
Greer 2015	Investigate indexical meanings of creaky voice and reasons for the increase in creaky voice use in young American female speakers	Speakers instructed to produce sentences in: modal, 50% modal-breathy, phrase final breathy, 50% modal-creaky, phrase final-creaky	Sentences	6 speakers; 3 male, 3 female	Listeners asked to decide which speaker sounds more masculine (for male speakers) or which sounds more feminine (for female speakers)	19 listeners; 12 female, 7 male	For male speakers, creakier voices tended to be perceived as more masculine compared with less creaky or non-creaky voice qualities. For female speakers, more breathy voices tended to be perceived as more feminine compared to less breathy or non-breathy voice qualities.
Holmberg et al. 2010	To examine the usefulness of phonetograms and aerodynamic measures for voice assessment of transgender women	Airflow used to measure air pressure and airflow rate	Read passage	25 Australian transgender women; 12 cisgender women & 12 cisgender men, likely also Australian but not specified	Self-reported estimate of passing as women (speakers); 100 point scale from 'very male' to 'very female' (listeners)	10 women and 10 men	Transgender women had low rates of breathy voice, and no relationship between breathy voice and listener ratings. Transgender women who reported themselves having a 'croaky' voice were rated as less female and reported lower satisfaction with their voice.

Table 1.3: Table summarising previous research looking at the contribution of voice quality to gender perception

Article	**Study purpose**	**Method of producing/measuring VQ**	**Stimuli characteristics**	**Speaker demographics**	**Speaker gender rating scale**	**Listener demographics**	**Results/conclusions**
King, Brown & McCrea 2012	Examine the characteristics that lead to gender being identified or misidentified in the voices of transgender women and cisgender women with low-pitch voices	jitter and shimmer	Read passage	21 transgender women and 9 cisgender women	Binary categorization of 'male' or 'female'; 7-point scale from 'extremely masculine' to 'extremely feminine'	20 listeners. 15 female and 5 male.	No evidence that jitter or shimmer influence gender perception
Lee 2016	Investigate perceptual judgements of creaky voice	Speakers instructed to produce sentences in modal voice and creaky voice	Sentences	1 male and 1 female speaker, both native speakers of American English.	5 point Likert Scale from strongly disagree to strongly agree for both masculine and feminine	53 participants; 38 female, 15 male.	Creaky voice rated less feminine and more masculine than modal in the speech of both male and female speakers.
Owen & Hancock 2010	Examine relationship between speaker-rated and listener-rated vocal femininity and acoustic measures of the voice	f0, noise to harmonics ratio (NHR), voice turbulence index (VTI), shimmer percent, jitter percent	Isolated vowels and read passages	20 transgender women, 5 cisgender women, 5 cisgender men. Language and variety unspecified, but likely to be American English.	1000-point scale from masculine to feminine (both self and listener ratings)	25 listeners. 12 male and 13 female. Native English speakers.	No relationship between acoustic measures of voice quality and either self or listener rated femininity, but this does not necessarily mean that there is no relationship between voice quality and vocal femininity

Table 1.3: Table summarising previous research looking at the contribution of voice quality to gender perception

Article	Study purpose	Method of producing/measuring VQ	Stimuli characteristics	Speaker demographics	Speaker gender rating scale	Listener demographics	Results/conclusions
Palmer, Dietsch & Searl 2012	To describe vocal fold activity among transgender women producing a feminine voice	Endoscopic and stroboscopic procedures to investigate vocal fold configurations, including glottal closure, which is related to breathy voice	Isolated vowels	Nine transgender women; language and variety unspecified, but likely to be American English	Self-reported estimate of passing as women (speakers); binary choice of 'male' or 'female' (listeners)	10 listeners of unspecified demographics	Most participants used incomplete glottal glosure and a posterior glottal gap, suggesting use of breathy voice; participants who used this vocal fold configuration were more likely to report passing in everyday lives
Porter 2012	Examine which acoustic measurements of voice quality influence listener perception of speaker gender	(H1-H2, H1-A1, H1-A3, F1 bandwidth), noise parameters (VTI, NHR, SPI), fundamental frequency (f0), and perturbation of the signal (jitter and shimmer)	Read passage, carrier phrases, sustained vowels	37 cisgender female speakers; 27 cisgender male speakers including 1 with puberphonia; 4 transgender women	Binary choice of 'male' or 'female'; 100-point scales of 'least masculine' to 'most masculine' when categorised as male and from 'least feminine' to 'most feminine' when categorised as female	'Twenty-six listeners, 23 female and 4 male'. North American English speakers.	Several parameters associated with breathy voice were correlated with increased perceptions of femininity - but the parameters that correlated were different in sustained vowels and carrier phrases.

Table 1.3: Table summarising previous research looking at the contribution of voice quality to gender perception

Article	**Study purpose**	**Method of producing/measuring VQ**	**Stimuli characteristics**	**Speaker demographics**	**Speaker gender rating scale**	**Listener demographics**	**Results/conclusions**
Skuk & Schweinberger 2014	To determine the relative importance of different acoustic parameters to gender perception	Spectrum level (formant bandwidths and global spectral tilt) and aperiodicity morphed on a continuum between a male and a female voice using TANDEM-STRAIGHT	Meaningless VCV syllables	Two male and two female native German speakers	Binary choice of 'male' or 'female'	Baseline: 12 listeners (6 male, 6 female); Experiment 1: 16 new listeners (8 male, 8 female); Experiment 2: 16 new listeners (8 male, 8 female)	Spectrum level (formant bandwidths and global spectral tilt) has an effect on perception of speaker gender, but aperiodicity does not
Van Borsel, Janssens & De Bodt 2009	Examine whether breathy voice influences perception of speaker femininity	Speakers instructed to produce 'normal' and breathy voice qualities	Isolated vowels	Twelve female speech and language pathologists or final year speech and language pathology students who were native speakers of Dutch	Experiment 1: five-point rating scale with 1 (little feminine) and 5 (very feminine); Experiment 2: listener asked to indicated which was the most feminine out of two samples	20 male and 20 female native Dutch speakers	Breathier samples judged to be more feminine

1.4 The present study

This study seeks to begin to resolve the issue of whether or not voice quality, here restricted to phonation, has a role in cueing gender perceptions and investigate the following main research questions:

- How does voice quality affect perceptions of speaker gender?
- How does voice quality interact with fundamental frequency when cuing gender perceptions?

I also hope to begin to disentangle what I see as being two potential confounds in previous work on the role of voice quality in gender perception: the sociolinguistic background of listeners, and the type of response used to measure 'perceived gender', and investigate the following secondary research questions:

- Does listeners' sociolinguistic background and exposure to culturally-specific gender differences in voice quality affect whether voice quality contributes to gender perception?
- How does the type of response used to measure 'perceived gender' affect whether voice quality contributes to gender perception?

With the exception of studies that seek to investigate the causes, and indexical meanings of, high rates of creaky voice used by American women, most previous research looking at the role of voice quality in gender perception has focused on the role of breathy voice. As some research on UK varieties of English has not only found increased breathiness in women, but increased creakiness in men, this study seeks to investigate whether creaky voice usage has any effect on gender perceptions. However, research investigating breathy voice has also found conflicting results, and there are also differences in breathy voice usage between different varieties of English, so this study will also investigate the role of breathy voice in gender perception.

To investigate these questions, I conducted two perception experiments. In these experiments, I used stimuli where the phonation type was either produced as, or manipulated to be, breathy, modal or creaky. Additionally, the overall f0 of stimuli was manipulated, so that the median lay at either a typical 'male' level of 120 Hz, 'androgynous' level of 165 Hz, or a typical 'female' level of 210 Hz. Listeners were asked to indicate whether they thought the speaker was a man or a woman and rate the masculinity, femininity and naturalness of scales of 1-100 using a slider.

In Experiment 1, I used Scottish English listeners from the central belt of Scotland. I expect the gender differences in voice quality in Scottish English to affect how Scottish listeners perceive the gender of the stimuli. Specifically, I predict that overall, breathier samples will be more likely to be perceived as feminine and spoken by a woman, while creakier samples will be more likely to be perceived as masculine and spoken by a man. Where pitch is ambiguous, I expect that having a creaky or breathy voice quality may alter the perception of whether a man or a woman is speaking. Where pitch falls in the average level of a cisgender man or woman, I expect that breathy voices will be perceived as more feminine and that creaky voices will be perceived as more masculine, but that pitch will take precedence over voice quality when listeners are asked to determine whether a man or a woman is speaking.

In Experiment 2, I used North American listeners to investigate the contribution of listeners' sociolinguistic background. If culturally-specific gender differences in voice quality production affect how listeners perceive voices, I expect that North American listeners may produce different responses to Scottish listeners. For example, as we know that in the US, young women tend to be creaky (Yuasa 2010), I expect American English listeners may perceive creaky voices as more feminine than Scottish listeners.

Additionally, in both experiments, I expect that there will be a slight overall bias for the stimuli to be perceived as feminine and spoken by a woman, as they are originally produced by a female speaker.

Chapter 2

Methods

In order to investigate how voice quality affects perceptions of speaker gender and how voice quality interacts with f0 when cueing gender perceptions, two perception experiments were carried out. The stimuli listeners heard were all initially produced by a female speaker and each stimulus was manipulated to have an average f0 of either 120 Hz, 165 Hz, or 210 Hz, and either produced with a modal, breathy or creaky voice quality, or manipulated to have a modal, breathy or creaky voice quality. In the two experiments, listeners were asked to listen to these stimuli and identify whether the voice sounded like that of a man or a woman, or rate how masculine and feminine the voices sounded on scales of 1-100. In Experiment 1, participants consisted of listeners from around the Central Belt of Scotland who had lived in the Greater Glasgow area for at least a year. In Experiment 2, participants consisted of listeners who were born in and currently resided in North America, who were recruited online using Prolific and completed the experiment online. This allowed investigation into whether the sociolinguistic background of listeners had an effect on how voice quality affected listener perception of speaker gender.

2.1 Stimuli

2.1.1 Comparison of techniques for creating stimuli with voice quality differences

There are two broad groups of approaches for creating stimuli with differences in voice quality to enable investigation into perception of voice quality. The first is using natural stimuli that have been produced with differences in voice quality, while the second uses synthetic stimuli that have been manipulated in some way to have different voice qualities.

Voice quality perception: Approaches using natural stimuli

Experiments that use natural stimuli have several different possible ways to gather their stimuli. As in Bishop and Keating (2012), stimuli can be gathered from a number of speakers who end up producing the stimuli with different voice qualities without being asked. While the naturalness of stimuli in experiments is certainly an advantage, there are also drawbacks to this approach: researchers do not have much control over what aspects of the voice vary between speakers, and the relationship between voice quality measures of the stimuli and listener judgements can often be difficult to interpret. For example, in Bishop and Keating (2012)'s study investigating whether voice quality cues allow listeners to identify whether a sound is high or low in the overall pitch range of that speaker indirectly through helping to identify speaker sex, the measurement of H2*-H4* was found to correlate with listener judgements of speaker sex as 'male'. However, as H2*-H4* is a relatively uncommon measure of voice quality, the authors had some difficulty in interpreting this result.

Using speakers trained in voice quality can help to mitigate some of these problems surrounding control of the stimuli and interpretation of the results. Van Borsel et al. (2009) asked speech and language therapists and speech and language therapy students to produce /a/ first with their habitual voice quality and then with a breathy voice. As these speakers were acquainted with variations in voice quality and had been trained in voice quality production as part of their speech and language therapy training, this allowed more control over the stimuli that was produced. Furthermore, Van Borsel et al. (2009) only retained samples where the pitch and intensity of the breathy productions fell within 5% of the production with habitual voice quality, giving them a further degree of control. Additionally, a panel of speech and language therapists to rated the breathiness of each sample; habitual samples produced with a breathy voice quality and breathy samples produced with a non-breathy voice quality were disregarded. This also avoided Bishop and Keating (2012)'s problem of connecting voice quality measures to perceptual voice quality categories. However, if longer samples than steady state vowels were used, there would be more scope for variability both within and between speakers, meaning it would be more difficult to retain this level of control with the same method.

Voice quality perception: Approaches using synthetic stimuli

According to source-filter theory, the speech signal can be separated into understood as consisting of a source, that is, the airflow through the glottis, and a filter, corresponding to the resonances in the oral and nasal cavities that shape the source (Fant 1960). Where experiments investigating voice quality perception use synthetic stimuli, they often use a technique called inverse filtering to separate the glottal source from the filter, then manipulate aspects of the source signal to alter the voice quality, before resynthesizing the original sound by reapplying the filter. The main advantage of this approach over using a natural voice is an increased degree of control over the stimuli. The inverse filtering can be done manually, as in Gobl and Ní Chasaide (2003), which allows for longer portions of speech to be inverse filtered more accurately than in automatic methods and produces a final output that is close to natural speech. However, the drawback of this is that it is extremely time consuming to undertake. Alternatively, software such as the UCLA Voice Synthesizer (Kreiman et al. 2016) allows highly accurate automatic inverse filtering of steady state vowels and produces an output that listeners cannot differentiate from natural speech (Gerratt and Kreiman 2001; Kreiman and Gerratt 2005). However, it is unable to process longer or more complex portions of speech.

The implementation of the Klatt synthesizer (D. Klatt and L. Klatt 1990) in Praat (Boersma and Weenink 2019), known as KlattGrid (Weenink 2009), allows automatic copy synthesis of a speech signal. In this method, oral formants are computed by LPC analysis, pitch is determined by autocorrelation method described in Boersma (1993), and then the intensity of any period noise in the signal is measured. The results of these analyses are then used to fill the pitch tier, oral formant grid and voicing amplitude tiers in a resulting KlattGrid where the time domain is the same as that of the original sound. With the pitch, intensity and oral formants now separated, the user can manipulate aspects of the glottal source, such as spectral tilt, before resynthesizing the speech signal.

The advantage of this approach is that is makes manipulation of aspects of the glottal source possible over longer stretches of speech that other inverse filtering methods, while being less time-consuming than interactive inverse filtering methods. However, in comparison to other methods of separating the glottal source from the filter, it produces a far less natural sounding output. In addition to this, KlattGrid also assumes constant, periodic voicing in the sound that it copy synthesizes, constraining the kinds of sounds that can be successfully copy synthesized using this method.

2.1.2 The present study: incorporating natural and synthesised stimuli

The original intention of this study was to use only synthesised stimuli, creating differences in voice quality with KlattGrid using the method discussed above. However, after recording the voices of two initial participants, the output of the Klatt synthesizer was decided to be too unnatural to use on its own. I therefore decided to record two new speakers, who were able to produce voice quality differences naturally, and use voice quality differences that were produced naturally by the speakers, in addition to producing synthesized differences in voice quality using KlattGrid from sentences recorded by the speakers in modal voice. However, due to difficulties with the male speaker in both the natural production of voice quality differences and KlattGrid synthesis based on his voice, I decided to proceed using only the female speaker's voice. See section below on Speaker characteristics for more details on the speaker selection process.

The present study, then, involves both stimuli with voice quality differences produced naturally by a female speaker, with synthesized stimuli with voice quality differences produced using KlattGrid, based on the same female speaker's voice.

2.1.3 Selection of sentences for stimuli

The stimuli consisted of a selection of phrases and sentences, such as 'We were running' or 'A long year'. The sounds that the stimuli could contain were constrained by the limitations of the automatic copy synthesis procedure in the Praat (Boersma and Weenink 2019) implementation of the Klatt synthesizer (D. Klatt and L. Klatt 1990), known as KlattGrid (Weenink 2009). Because it works by extracting the oral formants, pitch and intensity from the original sound, it assumes constant voicing and does not cope well with silence, voiceless sounds or frication. This means that it works best with steady state vowels, but it can also work well with approximants and liquids and can generate acceptable results with nasals. Voiced plosives, fricatives and affricates can also generate acceptable outputs but are less reliable. Therefore, 108 short phrases and sentences containing only vowels, approximants, liquids and nasals were selected (see Appendix B).

2.1.4 Speaker characteristics

Before it was decided to also use stimuli with naturally produced differences in voice quality, two cisgender speakers, one female (Speaker 1) and one male (Speaker 2), who were naive to the topic of the experiments, were recorded. Both originated from the Southside of Glasgow and currently resided in Glasgow. However, as the output of the KlattGrid synthesis based on their voices sounded rather unnatural, I decided to recruit two new speakers, one male and one female, who had the ability to produce differences in voice quality naturally. Speaker 3 was a male speaker originating from Glasgow who had received previous phonetic training to be able to produce differences in voice quality. However, on the day of recording he encountered difficulties in producing modal voice and creaky voice, and stimuli produced using KlattGrid synthesis based on his modal voice had a distinctly unnatural quality - this is presumed to be due to his overall whispery voice quality, creating aspiration noise that created difficulties with copy synthesis in KlattGrid because it did not meet the assumptions of the synthesizer.

The final stimuli were recorded by Speaker 4, a cisgender female speaker who responded to an email advertisement looking for actors who had the ability to control aspects of their voice with a Glasgow accent. The speaker was a theatre student who originated from and currently resided in the West End of Glasgow, Scotland. She reported having English as her first language, never having lived in another country for an extended period of time, and not having any kind of speech or hearing disorder to the best of her knowledge.

2.1.5 Recording procedure

The speaker first completed a short questionnaire to determine her eligibility and then recorded the stimuli, reading from a sheet of paper. The stimuli were recorded in a sound attenuated booth with a Sennheiser MKH 40 P 48 cardioid condenser microphone in Audacity (Dominic Mazzoni 2018) in monoaural sound at 44100Hz. The microphone was positioned approximately 30 centimetres away from the lips and the recording level was calibrated to peak at between -4 and -12 when speaker said /ɑ/ to avoid the recording being too quiet or clipping.

The speaker first listened to recordings of modal voice, creaky voice, breathy voice and whispery voice produced by Laver (1980). She was then asked to imitate the auditory quality of modal voice, creaky voice, and produce a ‘breathy’ voice quality with audible glottal friction that lay somewhere on the auditory continuum of whispery-breathy voice that felt natural to produce. She was allowed to practice until she felt confident that she was able to produce the target voice qualities and received pointers from the researcher if she strayed from the target voice quality. She was then instructed to read the sentences in the target voice qualities, to pause slightly between each sentence, to stay as still as possible and avoid touching or rustling the paper and to pause if she needed to turn the page or have a drink of water.

The sentences that had been selected as outlined in Section 2.1.2 were presented to the speaker on a sheet of paper, where the 108 sentences were presented in three columns under the headings ‘Breathy’, ‘Creaky’ and ‘Modal’, each containing 36 sentences. The speaker read through the entire selection of stimuli in modal voice and then down each column in the voice quality marked at the top of the column. The original intention was to the use 18 of the 36 stimuli produced in the target voice quality of each column as the stimuli with naturally produced voice quality differences, and 54 out of the 108 sentences produced in modal voice to create the stimuli with synthesized voice quality differences. However, due to varying levels of success in the output of the copy synthesis process, some of the resulting stimuli used sentences with the same lexical content.

2.1.6 Acoustic characteristics of natural stimuli

To ensure that the speaker managed to produce the target modal, breathy and creaky voice qualities, the acoustic characteristics of stimuli produced with naturally occurring voice quality differences were measured using VoiceSauce. (Shue 2010; Yen-Liang Shue 2011).

As we have seen, there are various problems with interpreting H1-H2 when it is used alone, as differences in H1-H2 are relative rather than absolute. Here I will follow guidance by Garellek (2019), of combining H1-H2 with Harmonics-to-Noise Ratio to characterise differences in voice quality. I will use the formant corrected measure of H1-H2, that is, H1*-H2*, and take the measurement of HNR in the cepstral domain between 0-500Hz, known as HNR05.

Measurements were taken in VoiceSauce across the whole sentence, which was possible because the sentences were designed to be voiced throughout and contain only vowels, approximants and nasals. As I was interested in the overall quality of the utterances, I did not take the prosodic structure of the utterances into account.The HNR05 and H1*-H2* measurements were taken every millisecond over the course of each stimulus.

If the speaker successfully produced the distinction betwwen modal, breathy and creaky voice, we can expect creaky voice to have a lower H1*-H2* and a lower HNR, relative to modal voice, while breathy voice should have a higher H1*-H2* and a lower HNR, relative to modal voice.

Each stimulus was coded for the intended voice quality produced and effects of intended voice quality on HNR05 and H1*-H2* were modelled in two simple linear regression models in R v.3.6.1 (R Core Team 2019).

Table 2.1 summarises the results of the linear regression models prediction H1*-H2* and HNR05 as a function of the voice quality that the speaker who recorded the stimuli was asked to produce.

As shown in Figure 2.1, relative to modal voice, breathy voice had a significantly higher H1*-H2* (β = 0.542, SE(β) = 0.063, t=8.548, p<0.001) while creaky voice had a significantly lower H1*-H2* (β = -1.302, SE(β) = 0.064, t=-20.509, p<0.001).

As shown in Figure 2.2, relative to modal voice, both breathy voice (β = -0.9611, SEβ = 0.1178, t=-8.156, p<0.001) and creaky voice displayed a significantly lower HNR05 (β = -8.0567, SEβ = 0.1180, t=-68.293, p<0.001).

Table 2.1: Summary of linear regression models predicting the H1*-H2* and HNR05 as a function of target voice quality. For each variable, estimate is shown first, followed by standard error in brackets

	Dependent variable:	
	H1*H2*	HNR05
	(1)	(2)
Constant	8.230***	35.574***
	(0.046)	(0.085)
Breathy voice	0.542***	−0.961***
	(0.063)	(0.118)
Creaky voice	−1.302***	−8.057***
	(0.064)	(0.118)
Observations	53,860	53,860
R^2	0.017	0.096
Adjusted R^2	0.017	0.096
Residual Std. Error (df = 53857)	5.966	11.084
F Statistic (df = 2; 53857)	464.197***	2,848.207***
Note:	*p<0.05; **p<0.01; ***p<0.001	

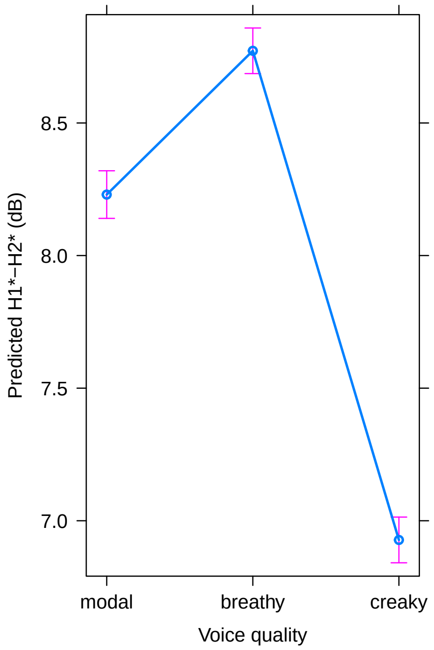


Figure 2.1: Effect of target voice quality on H1*-H2*

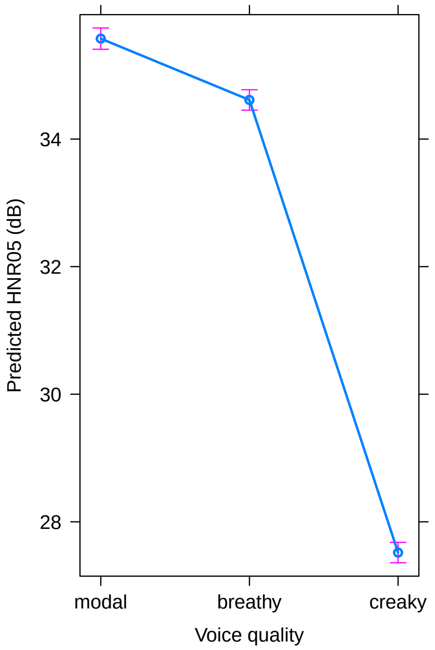


Figure 2.2: Effect of target voice quality on HNR05

This suggests that the speaker successfully produced a distinction between modal, breathy and creaky voice.

Table 2.2: KlattGrid manipulations made to produce voice quality differences in the stimuli

	Modal	**Breathy**	**Creaky**
Flutter	0.1	0.25	0.25
Open Phase	0.7	0.95	0.4
Spectral tilt	10	30	5
Aspiration noise	No change	Varied dynamically according to voicing amplitude	No change
F1 bandwidth	No change	Increase by 10%	No change
Double pulsing	None	None	None above median f0 0.125 when below the median f0 0.25 when below the 0.375 percentile of f0

Figure 2.1 and Figure 2.2 and were created using the effects package (Fox and Weisberg 2019) in R v.3.6.1 (R Core Team 2019) and show modelled values.

2.1.7 Creation of synthesized stimuli using KlattGrid

Manipulation of the stimuli was done using a Praat script, shown in Appendix A. The script first manipulated the median pitch to one of three levels: either 120Hz to represent a typical ‘male’ voice, 165Hz to represent an ‘ambiguous’ voice, or 210Hz to represent a typical ‘female’ voice.

Using copy synthesis, the resulting sound files were transformed into KlattGrids so that voice quality could be manipulated. The values for each parameter were chosen based on the values given in D. Klatt and L. Klatt (1990) and Gobl and Ní Chasaide (2003) for synthesis of different voice qualities, but changed where those values sounded unnatural or did not sufficiently create the impression of the desired voice quality.

To create the stimuli for the modal condition, flutter was set to 0.1, open phase was set to 0.7 and spectral tilt was set to 10. Other parameters were left at the values determined by the copy synthesis function or the default values. To create the stimuli for the breathy condition, flutter was set to 0.25, the bandwidth of the first formant was increased by 10%, open phase was set to 0.95, and spectral tilt was set to 30. Aspiration noise was varied dynamically according to the amplitude of voicing, so that the aspiration noise was equal to 40% of the amplitude of voicing.

To create the stimuli for the creaky condition, flutter was set to 0.25, open phase was set to 0.4 and spectral tilt was set to 5. Because in English creaky voice is more common in contexts where the f0 is low, double pulsing was set to vary dynamically according to f0, staying at 0 when the f0 lay above the median pitch, increasing to 12.5 when f0 dropped below the median, and increasing to 25 when f0 dropped below the first quartile.

The KlattGrid was then converted back to a sound file.

2.2 Experiment 1

2.2.1 Participants

Experiment 1 investigated the impact of voice quality on Scottish listeners' gender perceptions. 33 participants were were recruited through advertising to staff and students at the University of Glasgow and through word of mouth. Participants were told that they should meet the eligibility criteria of being at least 18 years of age, originating from the wider the Central Belt of Scotland, and living in the Greater Glasgow area for at least a year. Out of 33 participants, 32 met this criteria. One participant who originated from and currently resided in Ayrshire, but had not lived in the Greater Glasgow area for at least a year was also included included. Participants were aged between 19 and 45. All reported being native speakers of English, with one participant reporting also being a native speaker of Scots and another also being a native speaker of German. Of the 33 participants, 25 identified as the gender they were assigned at birth, and of these 9 were male and 16 were female. Of those who reported not identifying as the gender they were assigned at birth, there was one transgender female participant, and 7 who selected 'other' and wrote in their gender as one of the following: non-binary, agender, gender fluid, genderflux, masculine non-binary, non-binary (agender). All participants reported having normal hearing.

2.2.2 Procedure

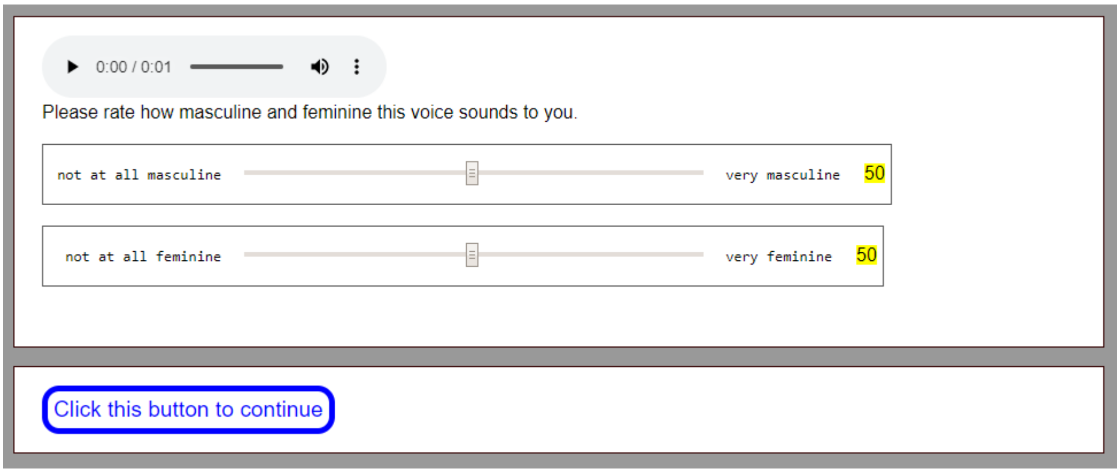
The study was created using PsyToolkit, an online tool for creating and running experiments online (Stoet 2010, 2017). The experiment was then conducted in-person on a desktop computer in a quiet room on the University of Glasgow campus. A copy of Experiment 1 can be accessed here: https://www.psytoolkit.org/cgi-bin/psy2.5.4/survey?s=ba7pc

Participants listened to stimuli through Sennheiser HD518 headphones.

First, participants completed a short questionnaire to collect some limited demographic information, then began the main portion of the experiment.

For each voice sample, participants were asked to rate how natural the sample sounds on a scale of 1-100 using a slider, and either:

- Indicate whether they think the speaker is a man, a woman, or neither/can't tell, as shown in Figure 2.4
- Or rate how masculine and feminine the voice sounds on scales of 1-100 using a slider, as shown in Figure 2.3.

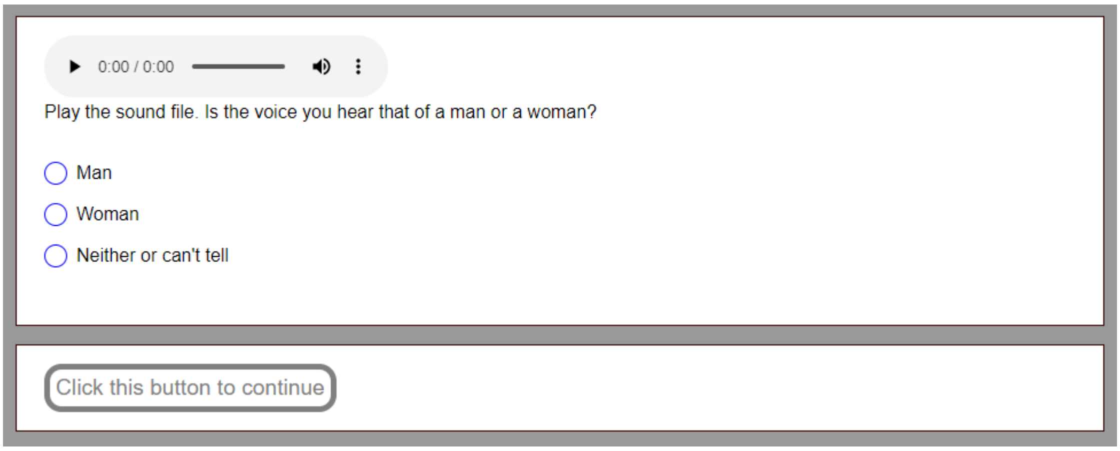


Please rate how masculine and feminine this voice sounds to you.

not at all masculine — very masculine 50

not at all feminine — very feminine 50

Figure 2.3: Screenshot from Experiment 1 showing listeners being asked to rate how masculine and feminine the voice they hear sounds to them



Play the sound file. Is the voice you hear that of a man or a woman?

○ Man

○ Woman

○ Neither or can't tell

Figure 2.4: Screenshot from Experiment 1 showing listeners being asked to indicate whether the voice they hear belongs to a man or a woman

The question on naturalness served as a distraction technique to encourage participants to believe that the experiment is focused on what makes synthetic speech sound natural, which was intended to explain why some of the stimuli sounded so unnatural and make participants feel more comfortable in making judgements about the gender of the stimuli.

At the end of the experiment, participants were asked to give optional feedback on their experience completing the experiment and what they paid attention to when making decisions about the gender of a speaker. Specifically, they were asked:

1. What did you pay attention to when deciding whether you thought the voices sounded like a man or a woman?

2. What did you pay attention to when rating how masculine/feminine a voice sounded?

3. Do you have any further comments about your experience completing this experiment?

Because of the distraction technique involved, at the end of the experiment participants were given the opportunity to leave their email address so that they could be contacted with a debrief as to the true purpose of the experiment after was complete.

2.2.3 Design

The independent variables were the naturalness of the stimuli (either natural or synthesized), the median f0 of the stimuli (210, 165, or 120Hz), and the voice quality (VQ) of the stimuli (modal, breathy, or creaky). The study involved a 2x3x3 within-subjects design where naturalness was crossed with voice quality and f0 to create 18 conditions. Within the condition of naturalness, the conditions are summarised in Table 2.3.

Participants heard a total of 108 stimuli, with 54 in the natural condition and 54 in the synthesized condition. Within each of these two conditions, participants heard each sub-condition six times in total, responding to each question on three occasions for each condition. The stimuli were presented in a randomized order.

The dependent variables were the listener ratings of perceived gender (man, woman, or neither/can't tell), perceived masculinity (rated on a scale of 1-100) and perceived femininity (rated on a scale of 1-100).

	Modal	**Breathy**	**Creaky**
210 Hz	Modal - 120	Modal - 165	Modal - 210
165 Hz	Breathy - 120	Breathy - 165	Breathy - 210
120 Hz	Creaky - 120	Creaky - 165	Creaky - 210

Table 2.3: Conditions of Experiment 1, where voice quality is crossed with median f0 (Hz)

2.3 Experiment 2

2.3.1 Participants

Experiment 2 investigated the impact of voice quality on non-Scottish listeners' gender perceptions. Participants were recruited using Prolific (www.prolific.co, 2014), an online tool for recruiting participants to take part in online experiments and compensate them for their time. Participants then took part in the experiment in the online version of Psytoolkit. Because a large amount of Prolific users who were native speakers of English were born in and currently resided in North America, participants were restricted to native English speakers born in and currently resident in the US or Canada. This was because if native English speakers from countries outside of North America were eligible, the sample would likely have been made up mainly of participants from the US or Canada and the number of participants from outside of North America would likely have been negligible.

40 participants took part in Experiment 2. All participants were born in and currently resided in the US or Canada and no participants reported having lived in the UK for a period of more than one month. All participants were native speakers of English and one participant participant reported being a native speaker of English and French. Participants were aged between 18 and 62. Of the 40 participants, there were 17 female participants and 23 male participants, including one transgender male. All participants reported having normal hearing.

Participants were screened to ensure that they were completing the experiment in a quiet space using headphones and a compatible web browser.

2.3.2 Procedure

The procedure of Experiment 2 was very similar to that of Experiment 1. There were slight differences in initial demographic questions and participants were given a debrief immediately after completing the experiments rather than being asked to leave their emails if they wanted to receive a debrief once data collection was complete, but the procedure of the main experiment was the same as in Experiment 1. A copy of Experiment

2 can be accessed here: https://www.psytoolkit.org/cgi-bin/psy2.5.4/survey?s=zsxzS.

Participants were allowed a maximum of 90 minutes to complete the experiment and took an average of 35 minutes.

2.3.3 Design

The design of Experiment 2 was the same as that of Experiment 1.

2.4 Statistical analysis

The results of Experiment 1 and Experiment 2 were analysed in R v.3.6.1 (R Core Team 2019). The two experiments were considered together to allow comparison between Scottish listeners from Experiment 1 and North American listeners from Experiment 2, with experiment/listener origin treated as an independent variable with two levels, Scottish (from the listeners of Experiment 1, recruited locally) and North American (from the listeners of Experiment 2, recruited online). Listener ratings of masculinity and femininity were analysed in two separate mixed-effect linear regression models, and listeners' categorical responses to the perceived gender of the speaker were analysed in a mixed-effect logistic regression model. Plots showing the effects of these models were created using the effects package (Fox and Weisberg 2019) and ggplot2 (Wickham 2016).

2.4.1 Analysis of masculinity and femininity ratings

Listeners rated the masculinity and femininity of each stimulus from 1-100 using a sliding scale. The full linear mixed-effects regression models were built using the `lmer()` function of the `lme4` package (Bates et al. 2015) with the `lmerTest` package (Kuznetsova et al. 2017) loaded to allow for model comparison using the `step()` function of the `stats` package (R Core Team 2019). The full models modelled masculinity and femininity as a function of the fixed effects of f0, VQ, origin and naturalness; the interactions between f0, VQ, and origin; and the random intercepts of sentence and participant.

The full models were then stepped down automatically using the `step()` function in R Core Team (2019) which adds and removes predictors and compares each model using Akaike's information criterion (AIC) in a Stepwise Algorithm to choose an optimal model.

The final model for masculinity, shown in (1), modelled masculinity as a function of of the fixed effects f0, VQ, origin, naturalness; the interaction between f0 and VQ; the interaction between f0 and origin; and the random intercepts of sentence and participant.

The final model for femininity, shown in (2), modelled femininity as a function of the fixed effects of f0, VQ, origin and naturalness; the interaction between f0 and origin; and the random intercepts of sentence and participant.

1. `lmer(masculinity ∼ VQ + origin + f0 + naturalness + (1 | sentence) + (1 | participant) + VQ:f0 + origin:f0)`
2. `lmer(femininity ∼ VQ + origin + f0 + naturalness + (1 | sentence) + (1 | participant) + origin:f0)`

2.4.2 Analysis of categorical gender responses

Listeners indicated whether they thought each voice belonged to a 'man', a 'woman' or 'neither/can't tell'. As it is not currently possible to run multinomial logistic regression with mixed effects in R, it was necessary to group together two of these responses into one factor. As the stimuli were all initially produced by a female speaker and expected to contain residual cues to this (for example, in the vowel formant frequencies) and thus elicit more 'woman' responses than 'man' or 'neither/can't tell', the 'man' and 'neither/can't tell' responses were grouped together into a 'not woman' category and a binomial generalized linear mixed-effects regression model was run using the `glmer()` function of the `lme4` package (Bates et al. 2015).

The full model modelled listener gender response as a function of the fixed effects of f0, VQ, origin and naturalness; the interactions between f0, VQ, and origin; and the random intercepts of sentence and participant.

As it is not currently possible to use the `step()` function to automatically step down a binomial generalized linear mixed-effects regression model in R, the full model was then stepped down manually removing predictors and comparing the AIC of each model in a log likelihood ratio test using `anova()`. This process is thus very similar to the `step()` function. First, the model was tested with the random intercepts removed, and they were both kept in the model as the test indicated that including the random intercepts improved the model. Then, the model was tested with each interaction removed, and the interaction between VQ and origin was removed as including it did not significantly improve the model, and other interactions were kept in as they did significantly improve the model. Finally, the model was tested with the fixed effects of naturalness removed, which was kept in as it improved the model. This resulted in a final model that modelled listener gender response as a function of f0, VQ, origin and naturalness; the interactions between voice quality, f0 and origin; and the random intercepts of sentence and participant.

- `glmer(response ~ VQ + origin + f0 + naturalness + (1 | sentence) + (1 | participant) + VQ:f0 + origin:f0)`

2.5 Qualitative analysis of participant feedback

At the end of the experiments, participants were asked the following questions:

1. What did you pay attention to when deciding whether you thought the voices sounded like a man or a woman?
2. What did you pay attention to when rating how masculine/feminine a voice sounded?
3. Do you have any further comments about your experience completing this experiment?

This qualitative data was analysed in order to examine whether it supported the quantitative findings of the experiment and investigate any additional cues to gender in the voices that listeners paid attention to.

2.5.1 Question 1 and Question 2: What did listeners pay attention to when...

Prior to analysis of the data, a list of initial codes was decided based on potential cues to speaker gender in the voice that have been identified in previous research. The initial codes were:

- Voice quality
- Pitch
- Intonation
- Vowel resonance
- Articulation

- Loudness

The data was coded manually by looking for words in each participant's answers that suggested they had paid attention to a particular cue. After a first pass of the data, I decided to include two additional codes, 'Tone/timbre' and 'Naturalness', due to multiple participants noting that they had paid attention to these. Additionally, 'Vowel resonance' was removed, as no participants noted that they had paid attention to this.

Finally, an additional code was added for participants mentioning that they had paid attention to what they thought other traits of the speaker were to help them identify the gender of the speaker.

Table 2.4 shows the final list of codes, along with examples from the data demonstrating the kinds of words participants used that suggested they were paying attention to that cue.

For some codes, it was relatively simple to identify the cue that the participant was referring to from the words they used, whereas for others, it was more difficult. For example, the language used to discuss pitch was fairly uniform between participants, with many using the word 'pitch' or terms like 'deep', 'high' or 'low' to describe voices. On the other hand, words that described voice quality were harder to categorise as they varied between different participants, and could have been intended as descriptions of other cues; for example, 'soft' was taken as a description of voice quality, but may have been intended to mean 'quiet' in some cases. On the other hand, words like 'harsh' and 'vocal fry' were straightforward to categorise as references to voice quality.

In addition, participants often demonstrated difficulty in pinpointing what exactly they had paid attention to, often borrowing words from music to describe what they had paid attention to in the voices. For example, the terms 'tone' and 'timbre', which were grouped together into a single code, may have been used by participants to refer to a range of cues, including pitch, vowel resonance, voice quality, intonation, and loudness.

2.5.2 Question 3: Additional feedback

There were no initial codes established for the analysis of participant's additional feedback.

After a first pass, the following codes were established:

- The experiment needed more male voices

Table 2.4: Codes used to code responses to Questions 1 and 2, and examples of the language that fell under each code

Code	**Examples**
Voice quality	‘Breathy-ness’, ‘soft’, ‘harshness’, ‘vocal fry’, ‘gruffer’, ‘nasality’, ‘raspiness’
Pitch	‘Pitch’, ‘deep or high pitch’, ‘High or low’, ‘depth of the voice’
Intonation	‘Lyrical’, ‘cadence’, ‘how much intonation there was’, ‘inflection’
Articulation	‘Clarity’, ‘sharpness/precision’, ‘the way words were pronounced’
Loudnesss	‘Volume’
Tone/timbre	‘Tone mainly’, ‘Timbre’
Naturalness	‘Electronic echo’, ‘less distorted’, ‘if they sounded human’
Other traits	‘Age of the speaker’, ‘like a female smoker’

- Comment on the clarity of the voices
- It was interesting or fun
- Comment about the rating scales

Chapter 3

Results

Table 3.1 presents the results of the statistical models analysing listener ratings of femininity, listener ratings masculinity, and listener categorisation of gender.

3.1 Listener perception of femininity

Listener ratings of femininity were analysed using a linear mixed effects regression model which predicted femininity as a function of the fixed effects of f0 (120Hz, 165Hz, 210Hz), VQ (modal, breathy, creaky), origin (Scottish participants from Experiment 1, North American participants from Experiment 2) and naturalness (natural, synthesized); the interaction between f0 and origin; and the random effects of sentence and participant.

Listener ratings of femininity showed variance according to participant (Variance = 137.16, SD = 11.711) and the lexical content of the sentence (Variance = 14.84, SD = 3.852).

Voice quality Figure 3.1 shows the effect of stimuli voice quality on listener perceptions of femininity.

Compared to modal voice stimuli, breathy stimuli were perceived as being significantly more feminine (β = 4.302, SE(β) = 1.170, t = 3.675, p <0.001). However, there was no significant difference in the perception of femininity between modal voice and creaky voice stimuli (β = 1.144, SE(β) = 1.290, t = 0.887).

Fundamental frequency level Compared to stimuli at 120Hz, stimuli at 165Hz (β = 23.935, SE(β) = 1.362, t = 17.577, p <0.001) and 210Hz (β = 33.570, SE(β) =

Table 3.1: Summary of mixed effects models predicting listener ratings of femininity, masculinity, and categorization of gender as a function of VQ, f0, listener origin/experiment, naturalness, and interactions between VQ and f0, as well as listener origin/experiment and f0. Reference levels are as follows: Modal voice (voice quality), 120 Hz (f0), Scotland (listener origin/experiment), natural (naturalness). For each variable, estimate is shown first, followed by standard error in brackets

	Dependent variable:		
	femininity	masculinity	response
	linear mixed-effects	*linear mixed-effects*	*generalized linear mixed-effects*
	(1)	(2)	(3)
Constant	59.163^{***} (2.446)	51.973^{***} (2.375)	1.218^{**} (0.406)
Breathy voice	4.302^{***} (1.170)	-5.891^{***} (1.585)	-1.630^{***} (0.319)
Creaky voice	1.144 (1.290)	-3.739^{*} (1.599)	-1.104^{**} (0.357)
165 Hz	23.935^{***} (1.362)	-32.247^{***} (1.753)	-5.018^{***} (0.434)
210 Hz	33.570^{***} (1.510)	-39.390^{***} (1.761)	-6.652^{***} (0.522)
North America	2.459 (2.909)	-16.560^{***} (2.846)	-1.323^{**} (0.437)
Synthesized	-20.405^{***} (0.841)	12.045^{***} (0.678)	2.657^{***} (0.184)
Breathy voice : 165 Hz		3.988 (2.204)	0.205 (0.507)
Creaky voice : 165 Hz		3.092 (2.239)	2.135^{***} (0.532)
Breathy voice : 210 Hz		5.766^{**} (2.170)	1.936^{**} (0.600)
Creaky voice : 210 Hz		1.198 (2.267)	1.023 (0.647)
165 Hz : North America	−1.002 (1.323)	9.316^{***} (1.345)	0.376 (0.261)
210 Hz: North America	-5.226^{***} (1.336)	13.832^{***} (1.351)	1.997^{***} (0.291)
Log Likelihood	−16,877.300	−16,911.760	−1,287.036
Note:			*p<0.05; **p<0.01; ***p<0.001

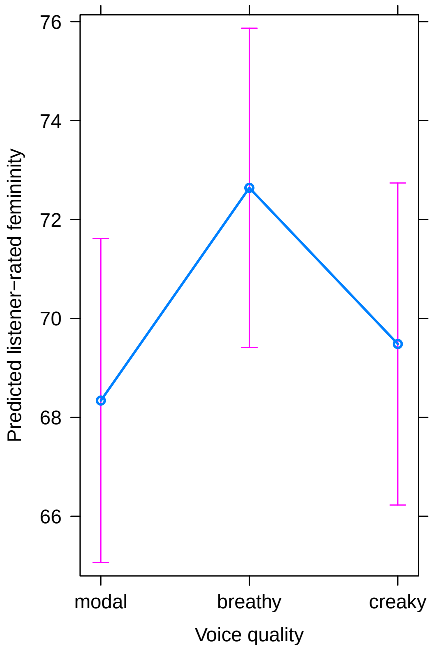


Figure 3.1: Effect of voice quality on listener ratings of femininity

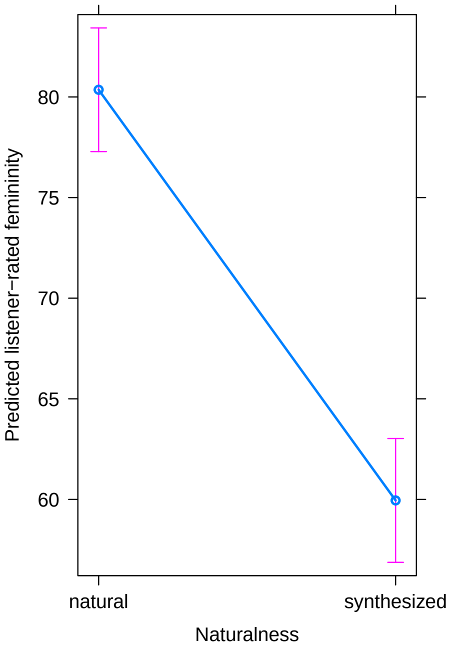


Figure 3.2: Effect of naturalness on listener ratings of femininity

1.510, t = 22.233, p <0.001) were rated as being significantly more feminine.

Naturalness Figure 3.2 shows the effect of naturalness on listener ratings of femininity.

Compared to the stimuli that were produced naturally, synthesized stimuli were perceived as significantly less feminine (β = -20.4053, SE(β) = 0.8411, t = -24.259, p <0.001).

Interaction of f0 level and experiment/listener origin Figure 3.3 shows the effect of fundamental frequency level and experiment/listener origin on listener ratings of femininity.

Scottish listeners from Experiment 1 and North American listeners from Experiment 2 did not show a significant difference in how feminine they rated stimuli overall. However, at 210Hz, North American listeners rated stimuli as being significantly less feminine than Scottish listeners did (β = -5.2258, SE(β) = 1.3357 , t = -3.912, p <0.001). There was no significant difference between Scottish and North American listeners' ratings of femininity at 165Hz (β = -1.0019, SE(β) = 1.3233 , t = -0.757).

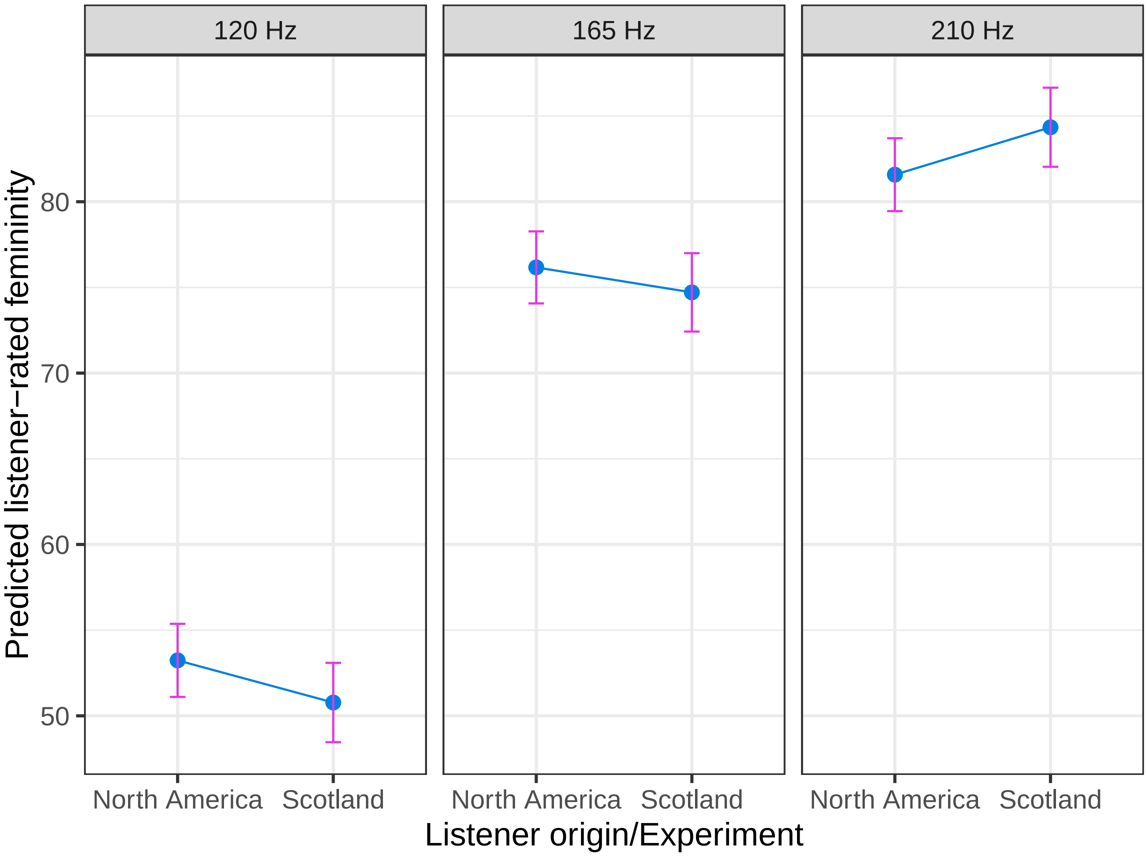


Figure 3.3: Effect of fundamental frequency level and experiment/listener origin on listener ratings of femininity

3.2 Listener perception of masculinity

Listener ratings of masculinity were analysed using a linear mixed effects regression model which predicted masculinity as a function of the fixed effects of f0 (120Hz, 165Hz, 210Hz), VQ (modal, breathy, creaky), origin (Scottish participants from Experiment 1, North American participants from Experiment 2) and naturalness (natural, synthesized); the interaction between f0 and origin; the interaction between f0 and VQ; and the random effects of sentence and participant.

Listener ratings of masculinity showed variance according to participant (Variance = 130.074, SD = 11.405) and the lexical content of the sentence (Variance = 3.973, SD = 1.993).

Voice quality Compared to modal voice stimuli, both breathy stimuli (β = -5.891, SE(β) = 1.585, t = -3.717, p <0.001) and creaky stimuli (β = -3.739, SE(β) = 1.599, t = -2.339, p <0.05) were perceived as being significantly less masculine.

Fundamental frequency level Compared to stimuli at 120Hz, stimuli at 165Hz (β = -32.247, SE(β) = 1.754, t = -18.390, p <0.001) and 210Hz (β = -39.3902, SE(β) = 1.761, t = -22.374, p <0.001) were rated as being significantly less masculine.

Experiment/listener origin Compared to Scottish listeners from Experiment 1, North American listeners from Experiment 2 perceived stimuli as being significantly less masculine (β = -16.560, SE(β) = 2.846, t = -5.819, p <0.001).

Naturalness Figure 3.4 shows the effect of naturalness on listener ratings of masculinity.

Compared to the stimuli that were produced naturally, synthesized stimuli were perceived as significantly more masculine (β = 12.045, SE(β) = 0.678, t = 17.757, p <0.001).

Interaction of f0 level and voice quality Figure 3.5 shows the effect of fundamental frequency level and voice quality on listener ratings of masculinity.

Compared to how breathy stimuli were rated in relation to modal stimuli at 120 Hz, breathy stimuli were rated as being more masculine relative to modal stimuli at 210 Hz (β = 5.766, SE(β) = 2.170, t = 2.657, p <0.01).

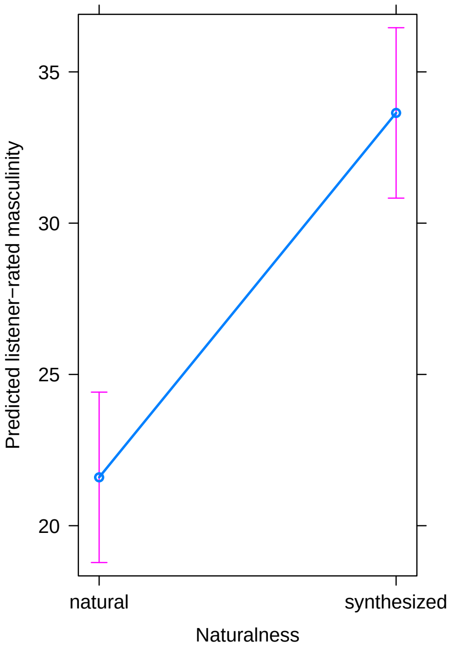


Figure 3.4: Effect of naturalness on listener ratings of masculinity

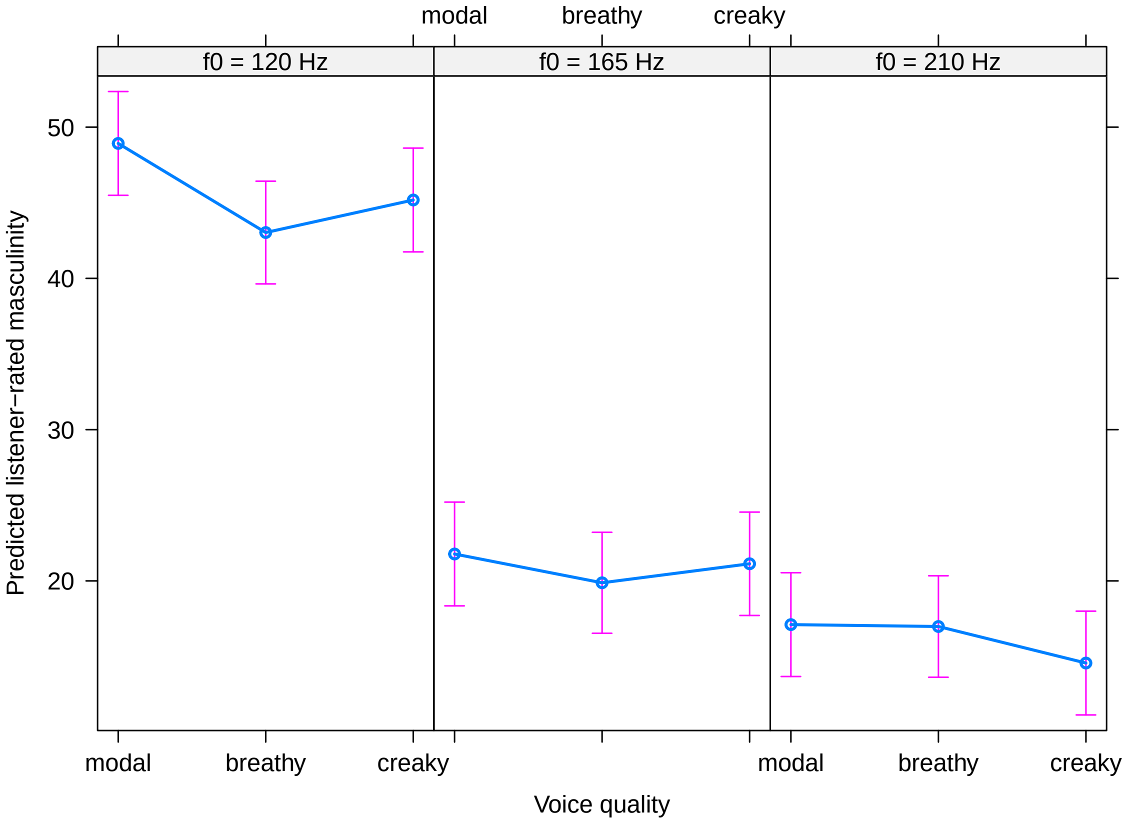


Figure 3.5: Effect of fundamental frequency level and voice quality on listener ratings of masculinity

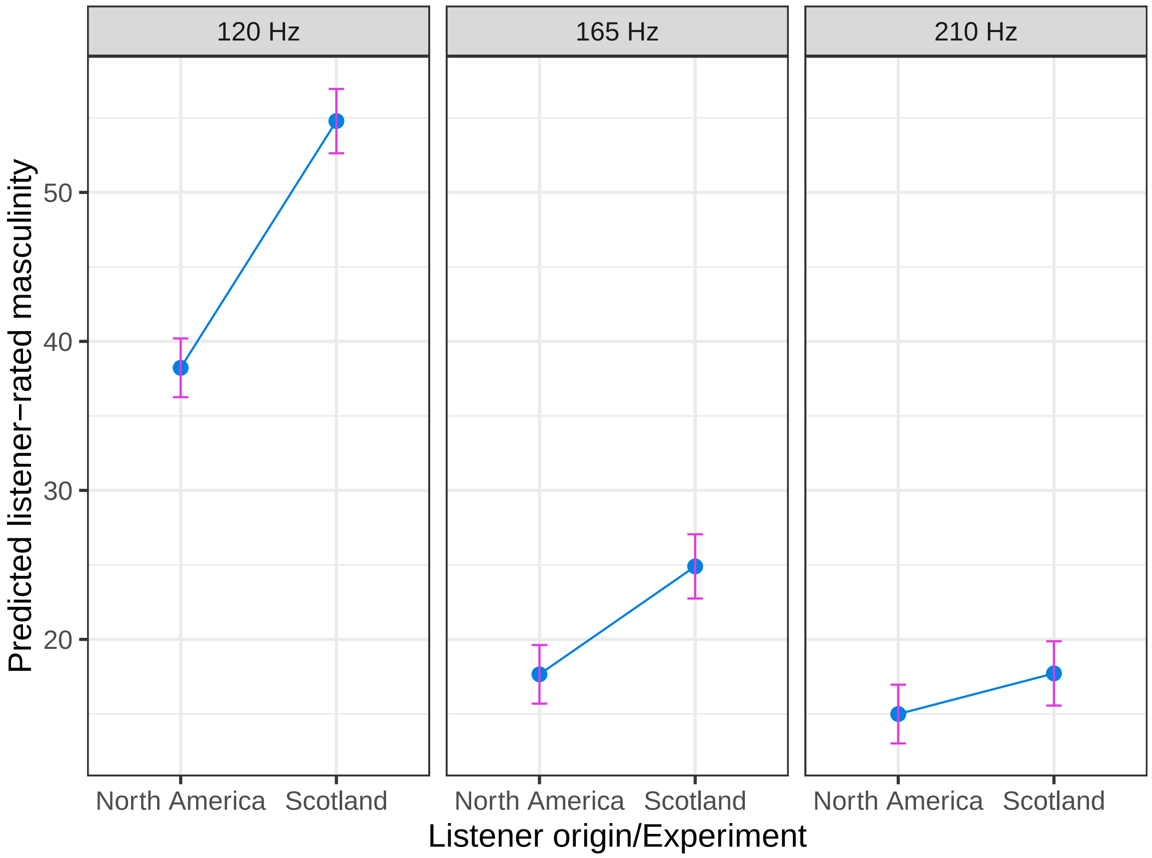


Figure 3.6: Effect of fundamental frequency level and experiment/listener origin on listener ratings of masculinity

Interaction of f0 level and listener origin Figure 3.6 shows the effect of fundamental frequency level and experiment/listener origin on listener ratings of masculinity.

Compared to how North American listeners rated stimuli in relation to Scottish listeners at 210 Hz, North American listeners rated stimuli at 165 Hz (β = 9.316, SE(β) = 2.170, t = 6.925, p <0.001) and at 210 Hz as more masculine (β = 13.8316, SE(β) = 1.3515, t = 10.234, p <0.001).

3.3 Gender responses

Listener's categorical responses of stimuli as being spoken by a woman or not were analysed using a generalized linear mixed effects regression model which predicted femininity as a function of the fixed effects of VQ (modal, breathy, creaky) and naturalness (natural, synthesized); the interaction between f0 (120 Hz, 165 Hz, 210 Hz) and origin (Scottish participants from Experiment 1, North American participants from Experiment 2); the interaction between VQ and origin; the interaction between f0 and VQ; and the random effects of sentence and participant.

Listener's categorisations of speaker gender as woman or not showed variance ac-

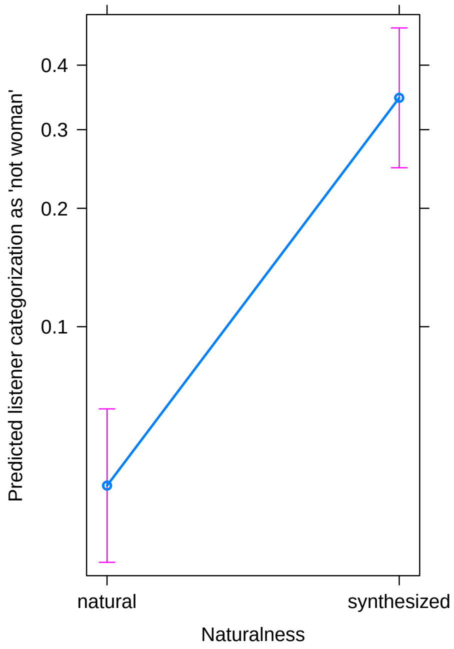


Figure 3.7: Effect of naturalness on listener categorisations of speaker gender identity

cording to participant (Variance = 2.975, SD = 1.725) and the lexical content of the sentence (Variance = 0.4136, SD = 0.6431).

Voice quality Compared to modal stimuli, both breathy stimuli (β = -1.630, SE(β) = 0.319, z = -5.107, p <0.001) and creaky stimuli (β = -1.104, SE(β) = 0.357, z = -3.091, p <0.01) less likely to be categorised as not being produced by a woman.

Fundamental frequency level Compared to 120 Hz stimuli, stimuli at both 165 Hz (β = -5.0178, SE(β) = 0.4339, z = -11.565, p <0.001) and 210 Hz (β = -6.652, SE(β) = 0.522, z = -12.746, p <0.001) were less likely to be categorised as not being produced by a woman.

Experiment/listener origin Compared to Scottish listeners from Experiment 1, North American listeners from Experiment 2 were less likely to categorise stimuli as not being produced by a woman (β = -1.323, SE(β) = 0.437, z = -3.023, p <0.01).

Naturalness Figure 3.7 shows the effect of naturalness on listener categorisations of speaker gender identity.

Compared to naturally produced stimuli, synthesised stimuli were more likely to

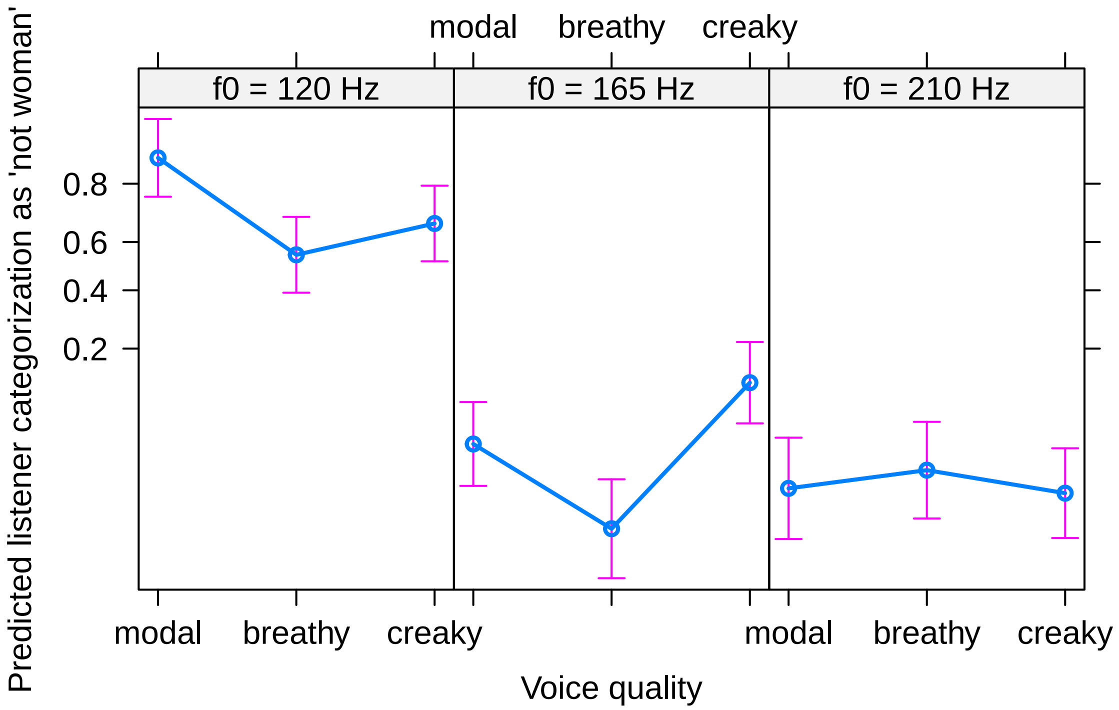


Figure 3.8: Effect of fundamental frequency level and voice quality on listener categorisations of speaker gender identity

be categorised as not being produced by a woman (β = 2.657, SE(β) = 0.1843, z = 14.414, p <0.001).

Interaction of f0 level and voice quality Figure 3.8 shows the effect of fundamental frequency level and voice quality on listener categorisations of speaker gender identity.

Compared to how creaky stimuli were perceived in relation to modal stimuli at 120 Hz, at 165 Hz, creaky stimuli were more likely to be perceived as not being produced by a woman in relation to modal stimuli (β = 2.135, SE(β) = 0.532, z = 4.015, p <0.001).

Compared to how breathy stimuli were perceived in relation to modal stimuli at 120 Hz, at 210 Hz, breathy stimuli were more likely to be perceived as not being produced by a woman in relation to modal stimuli (β = 1.936, SE(β) = 0.600, z = 3.228, p <0.01).

Interaction of f0 level and listener origin Figure 3.9 shows the effect of fundamental frequency level and experiment/listener origin on listener categorisations of speaker gender identity.

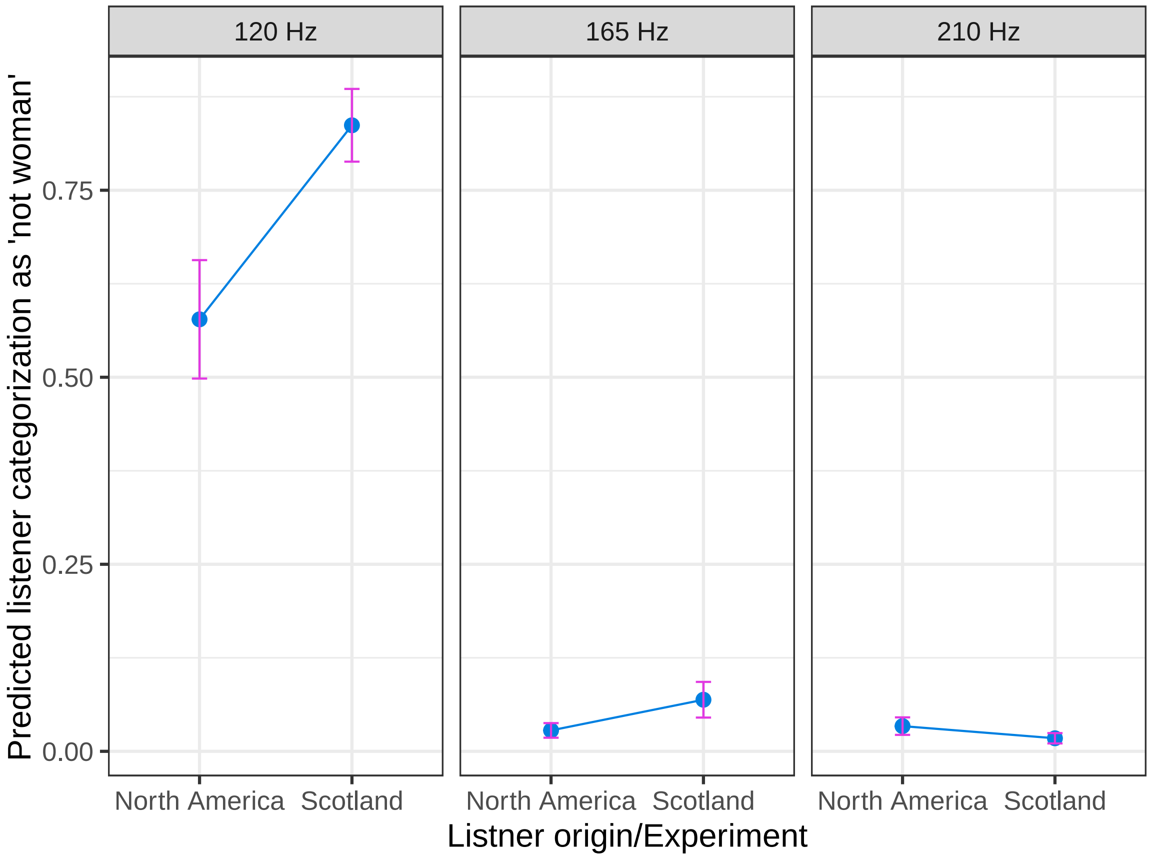


Figure 3.9: Effect of fundamental frequency level and experiment/listener origin on listener categorisations of speaker gender identity

Compared to how Scottish listeners perceived stimuli at 210 Hz in relation to how they perceived stimuli at 120 Hz, North American listeners were more likely to perceive stimuli at 210 Hz as not being produced by a woman (β = 1.997, SE(β) = 0.2910, z = 6.863, p <0.01).

3.4 Overview of qualitative data

At the end of the experiments, participants were given the option of providing feedback by responding to the following questions.

1. What did you pay attention to when deciding whether you thought the voices sounded like a man or a woman?

2. What did you pay attention to when rating how masculine/feminine a voice sounded?

3. Do you have any further comments about your experience completing this experiment?

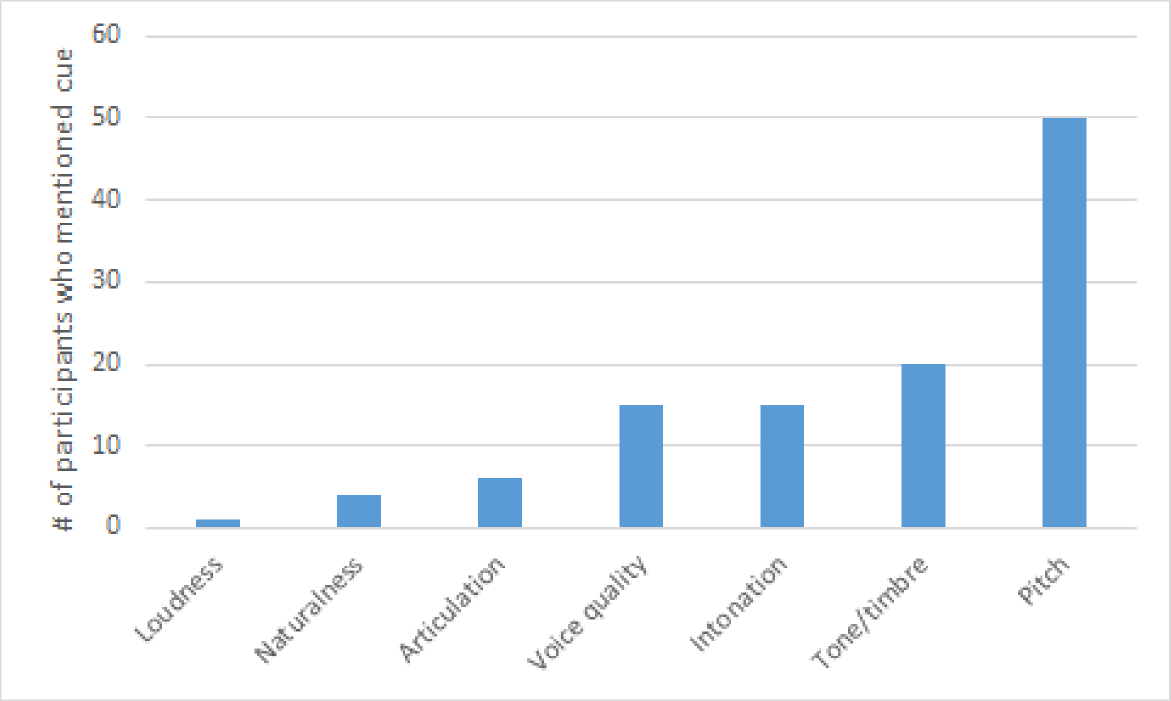


Figure 3.10: Cues identified in participant responses to Question 1

The response rate to these questions overall was relatively high. In Experiment 1, all 33 participants responded to Question 1, 31 out of 33 to Question 2, and 10 out of 33 provided further comments. In Experiment 2, 39 out of 40 participants responded to Question 1, 38 out of 40 responded to Question 2, and 17 out of 40 provided further comments.

3.4.1 Question 1 and 2

Figure 3.10 shows the cues that participants reported to paying attention to when deciding whether they thought the voices sounded like a man or a woman. In Question 1, pitch was the most recognised cue, being noted by 50 participants; followed by tone/timbre, which were noted by 20 participants; intonation and voice quality, which were each noted by 15 participants; articulation, which was noted by 6; naturalness which was noted by 4; and loudness, which was noted by 1. In addition to these cues, 4 participants mentioned that they had paid attention to other speaker traits when categorising the speaker's gender. These other traits were:

- Sexuality
- Accent
- Age

Figure 3.10 shows the cues that participants reported to paying attention to when

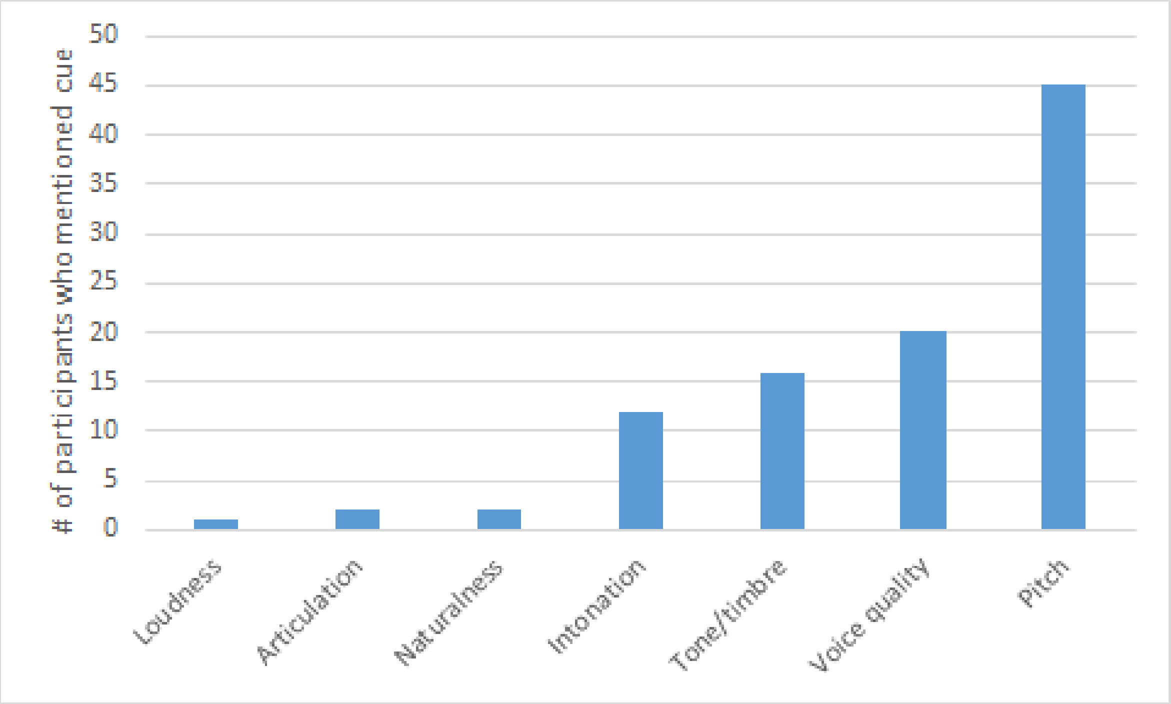


Figure 3.11: Cues identified in participant responses to Question 2

rating how masculine or feminine a voice sounded. In Question 2, pitch was again the most recognised cue, being noted by 45 participants; this was followed by voice quality, noted by 20; tone/timbre, noted by 16; intonation, noted by 12; naturalness and articulation, each noted by 2; and loudness noted by 1 participant.

In addition to these cues, in their responses to Question 2, 2 participants mentioned that they had paid attention to other speaker traits when rating how masculine or feminine a voice sounded. These other traits were:

- Age
- Being a smoker

3.4.2 Question 3

Question 3 asked participants for further feedback on their experience completing the experiment. 11 participants commented on the study being interesting or fun. 7 participants commented on the rating scales that were used. 3 participants commented that the experiment should have used more male voices. 5 participants commented on the clarity of the voices.

Chapter 4

Discussion

4.1 Discussion by individual results

4.1.1 Voice quality

Summary of predictions and findings Overall, I predicted that while breathy voice would be perceived as more feminine, less masculine, and more likely to be perceived as a woman speaking when compared with modal voice, creaky voice would be perceived as more masculine, less feminine, and less likely to be spoken by a woman.

Consistent with these predictions, when compared to how they rated modal voice, listeners rated breathy voices as more feminine and less masculine, and were more likely to categorise breathy voices as spoken by a woman over categorising them as a man or as 'neither/can't tell'. However, in contrast with these predictions, listeners rated creaky voices as less masculine than modal voices and similar in terms of femininity, and were more likely to categorise creaky voices as spoken by a woman over categorising them as a man or as 'neither/can't tell'.

Breathy voice The finding of breathy voice increasing listeners' perception of femininity relative to modal voice is broadly consistent with previous research (Addington 1968; Andrews and Schmidt 1997; Gorham-Rowan and Morris 2006; Greer 2015; Porter 2012; Van Borsel et al. 2009). It contradicts some previous research by Booz and Ferguson (2016) and Owen and Hancock (2010) that did not find a relationship between acoustic measures of voice quality that suggest increased breathiness. However, both Owen and Hancock (2010) and Booz and Ferguson (2016) suggest that their finding may have been partly due to the fact that the participants in their sample were relatively homogeneous in terms of their voice quality. As the present study used synthesized differences in voice quality, as well as elicited differences in voice quality from the speaker that formed distinct categories in terms of H1*-H2* and HNR, this ensured that the stimuli listeners heard did in fact display different voice qualities. This difference between breathy voice and modal voice in the stimuli in the present study may explain why breathy voice increased listener ratings of femininity, and decreased listener ratings of masculinity.

Similarly, the finding that breathy voice leads listeners to categorise the voice as being a woman speaking is also broadly consistent with previous research by Bishop and Keating (2012), Palmer et al. (2012), Porter (2012), and Skuk and Schweinberger (2014). To an extent, this contradicts the findings of King et al. (2012); however, King et al. (2012) only investigated the relationship between jitter and shimmer to listener ratings of speaker gender, which have been shown to have little relationship with listener perception of voice quality (Kreiman and Gerratt 2005).

This also contradicts findings of Holmberg et al. (2010), who found that breathiness had no relationship with gender ratings; however, Holmberg et al. (2010) found relatively low rates of auditorily-rated breathiness in their sample, suggesting that perhaps, similar to the situations in Booz and Ferguson (2016) and Owen and Hancock (2010), variation in voice quality in Holmberg et al. (2010)'s sample was too low for listeners to be able to use it as a cue to gender.

Creaky voice The finding that creaky voice is perceived as less masculine, but not significantly more feminine, than modal voice is contrary to predictions that creaky voice would be perceived as more masculine overall.

This prediction was in part based on production data from Scottish English suggesting that creaky voice is more common among male speakers. However, research conducted in London (Szakay and Torgersen 2015, 2019) suggests that usage of creaky voice is increasing among young female speakers. If this trend has spread to Scotland, this may explain why creaky voice was perceived as less masculine than modal voice, at least by Scottish listeners.

Furthermore, listeners in Experiment 2 originated from the US, and had limited exposure to Scottish English. Previous research on production data from American English suggests that increasingly, creaky voice is used more frequently by female speakers than male speakers in the US (Abdelli-Beruh et al. 2014; Syrdal 1996; Wolk et al. 2012; Yuasa 2010). In itself, this would help to explain the finding that creaky voice was perceived as less masculine; however, this finding actually contradicts work on the perception of creaky voice in US English, which finds that creaky voice is perceived as more masculine than modal voice in the speech of male speakers (Greer 2015), or in the speech of both male and female speakers (Lee 2016). Both Greer (2015) and Lee (2016) connect this perception of creaky voice as masculine with the low f0 of creak, suggesting that because male speakers tend to use a lower f0 in non-creaky speech, a lower f0 then becomes associated with masculinity and leads creaky voice to be perceived as masculine. Potentially, the fact that a female speaker was used to create the stimuli in the present study may have influenced whether creaky voice was perceived as masculine.

I also expected that voice quality would have a stronger effect on participant's perception of gender attributes than on their perception of speaker gender identity, altering perception of gender attributes at all f0 levels, but speaker gender identity only at an ambiguous level of 165 Hz. With this is mind, it is interesting to note that creaky voice had an effect on the perception of both masculinity and speaker gender identity, but not femininity. This could be due to interactions between the effect of voice quality and other factors like f0, listener origin and naturalness, which we will discuss below.

4.1.2 Fundamental frequency

The overall f0 of the stimuli was manipulated so that the resulting median f0 lay at 120Hz, 165Hz, or 210Hz. This manipulation of f0 was expected to serve as an important cue to speaker gender, cueing the listener to be more likely to hear a female voice at 210Hz and more likely to hear a male voice at 120 Hz. 165Hz was chosen as an ambiguous f0 level, where the impact of voice quality on the perception of gender could be maximised, based on previous research that finds it to be an ambiguous f0 range for transgender women Gelfer and Bennett (2013) as well as cisgender women with low voices (King et al. 2012).

These expectations were generally met, with stimuli in the 165 Hz and 210 Hz conditions being perceived as more feminine, less masculine, and as more likely to be a woman speaking than stimuli at 120 Hz. However, 120 Hz did not serve as a 'typically male' f0 range, as it was intended. In fact, as shown in Figure 4.1, at 120 Hz, just 31% of stimuli were categorised as being a 'man', while 33% were categorised as 'neither/can't tell', and 37% were categorised as being a 'woman'. Furthermore, at 165Hz, 84% of stimuli were categorised as 'woman'. This indicates that 120Hz, rather than 165 Hz, served as the more ambiguous f0 level for this speaker.

This is unsurprising for several reasons. First of all, the stimuli were not manipulated in terms of formant frequencies. As we saw in Section 1.1.3, formant frequencies have been found to be an important cue to speaker gender in previous research, in some cases being found to have an even bigger impact than fundamental frequency (Gelfer and Bennett 2013; Pisanski and Rendall 2011). In future research, it would be interesting to manipulate both fundamental frequency and formant frequencies to-

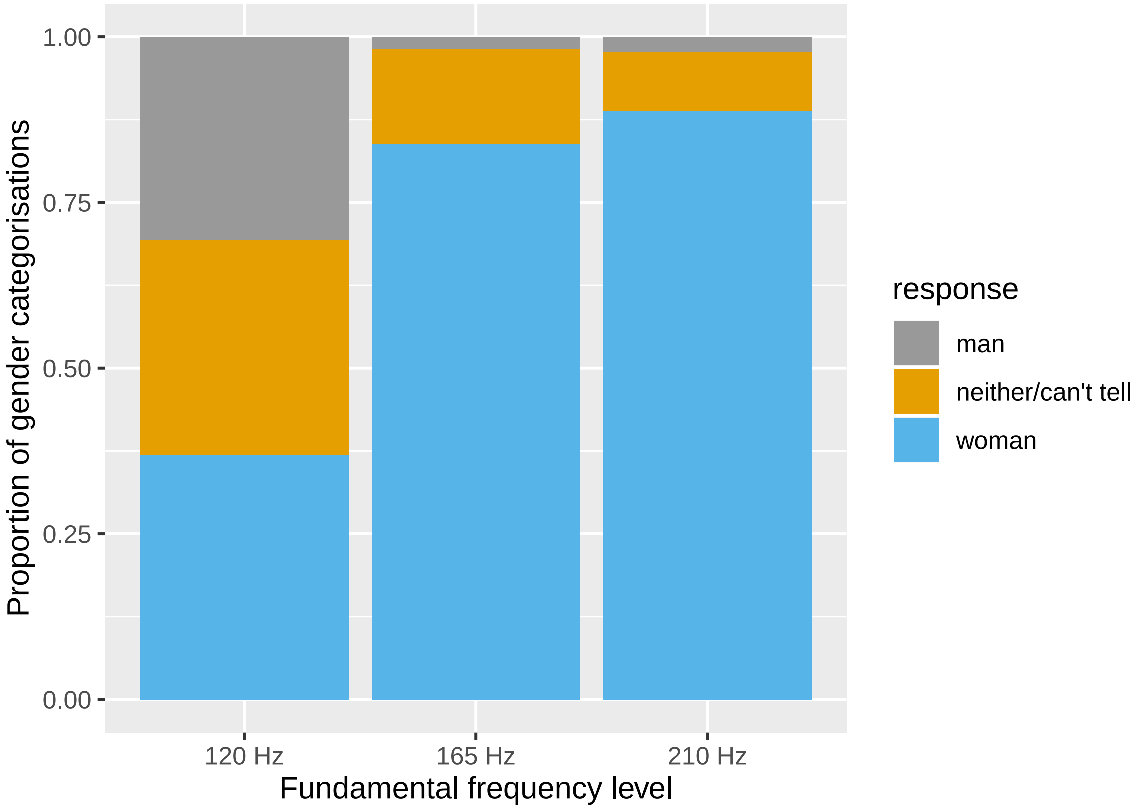


Figure 4.1: Overall distribution of gender responses by fundamental frequency level

gether to create a more ambiguous-sounding mid-point and maximise the influence of voice quality.

Furthermore, the original intention of this study was to have both a male and a female speaker to counterbalance the effect of the residual cues such as formant frequencies. However, I was unable to find a man who could create the required differences in voice quality or whose voice worked well with the KlattGrid copy synthesis, so only used a female speaker. This probably increased the impact that the residual cues to gender such as formant frequencies had on listeners' perceptions.

Additionally, while 165Hz sits in the middle of 120 Hz and 210 Hz in terms of Hertz, it is closer to 210 Hz than 120Hz in terms of semitones, and thus closer perceptually to the median f0 chosen to represent a 'typical female' voice than the f0 chose to represent a 'typical male' voice. In future research, it might be useful to choose have the median f0 chosen for the 'ambiguous' level be the same distance away from the 'typical male' and 'typical female' levels in terms of semitones, or for the researcher to conduct a norming study to evaluate what level of f0 sounds the most ambiguous to listeners.

4.1.3 Origin

I predicted that listeners' origin and domicile might affect how they rate the gender attributes and categorise the gender of the speech that they hear, due to culturally-specific gender differences in speech that they are used to hearing in their everyday lives. However, I did not have any specific predictions about what this difference might be.

Consistent with the expectation that they might differ, Scottish listeners from Experiment 1 differed from North American listeners from Experiment 2 in how they rated gender attributes and categorised the gender of the voices they heard. Specifically, North American listeners rated stimuli as sounding less masculine, though not more feminine, than Scottish listeners, and were more likely to rate stimuli as being spoken by a woman.

Eckert (2008) outlines how released [t] in American English can carry various meanings including 'British', 'prissy' and 'gay diva', through its association with clear speech. Although the stimuli in the present study did not contain any examples of released [t], it is possible that this association between British speech and various speech styles associated with an absence of masculinity in men persists into other features of British English. This may therefore help to explain the difference in ratings of masculinity and categorisation of gender identity between Scottish and North American listeners.

Additionally, it is possible that residual cues to speaker gender, in the form of vowel formant frequencies, may be more salient to North American listeners.

However, it is important to note that the listener origin variable does not only encompass whether a listener is Scottish or North American, but also whether they have taken part in the experiment online, or in person in a controlled environment. Although listeners from Experiment 2 were asked to confirm that they were using headphones and in a quiet place before taking part in the experiment, the type of headphones they used were not controlled for, and perhaps if listeners were using low-quality headphones this may have affected how they perceived the stimuli, and led them to rate stimuli as less masculine and more likely to be a woman speaking.

Additionally, there is a clear interaction between listener origin and perception of f0, which is discussed further below.

4.1.4 Naturalness

I expected that overall, synthesized voices would be perceived as less feminine, more masculine and less likely to be spoken by a woman than more natural stimuli that only received manipulation in terms of median f0. Consistent with this, stimuli that had been synthesized using KlattGrid were perceived as less feminine, more masculine and less likely to be spoken by a woman than natural stimuli. This is consistent with previous research (e.g. Coleman 1976) that has found that speech with a synthetic quality is less likely to be perceived as being spoken by a female speaker.

4.1.5 The interaction of voice quality and fundamental frequency

Summary of predictions and findings I expected that at levels of fundamental frequency that fell within the typical range of a cisgender male (120 Hz) or female speakers (210 Hz), differences in voice quality would only influence how listeners perceived the gender attributes of the speaker (i.e., masculinity and femininity), but not the speakers' gender itself. However, at an ambiguous f0 level (165 Hz), I expected that voice quality would have an effect on listeners' perception of speaker gender as well as on listeners' perception of speaker attributes. However, I found that voice quality had an effect, not only on perception of gender attributes, but also on perception of speaker gender identity at all levels of f0.

No interaction of fundamental frequency and voice quality emerged in the model of listener ratings of femininity, suggesting that the effect is consistent across all levels of f0 - that is, that breathy voice is perceived as more feminine than modal voice, but that there is no difference in listener ratings between modal and creaky voice. This is consistent with the prediction that breathy voice would be perceived as more feminine overall.

However, there were interactions between voice quality and f0 in terms of both masculinity and gender categorisations. In terms of masculinity, at the level of 210Hz, breathy voices were perceived as significantly more masculine relative to how breathy voices were rated in relation to modal stimuli at 120 Hz. In terms of categorisation of speaker gender identity, at 165 Hz, creaky stimuli were less likely to be perceived as being produced by a woman, relative to how creaky stimuli were categorised in relation to modal stimuli at 120 Hz. Furthermore, breathy stimuli were less likely to be perceived as being produced by a woman at 210 Hz, relative to breathy stimuli were categorised in relation to modal stimuli at 120 Hz.

Creaky voice at 165Hz The finding that creaky stimuli were less likely to be perceived as being produced by a woman at 165 Hz, relative to how creaky stimuli were perceived in relation to modal stimuli at 120 Hz, is consistent with the prediction that at an ambiguous f0 level, creaky voice might lead listeners to be more likely to

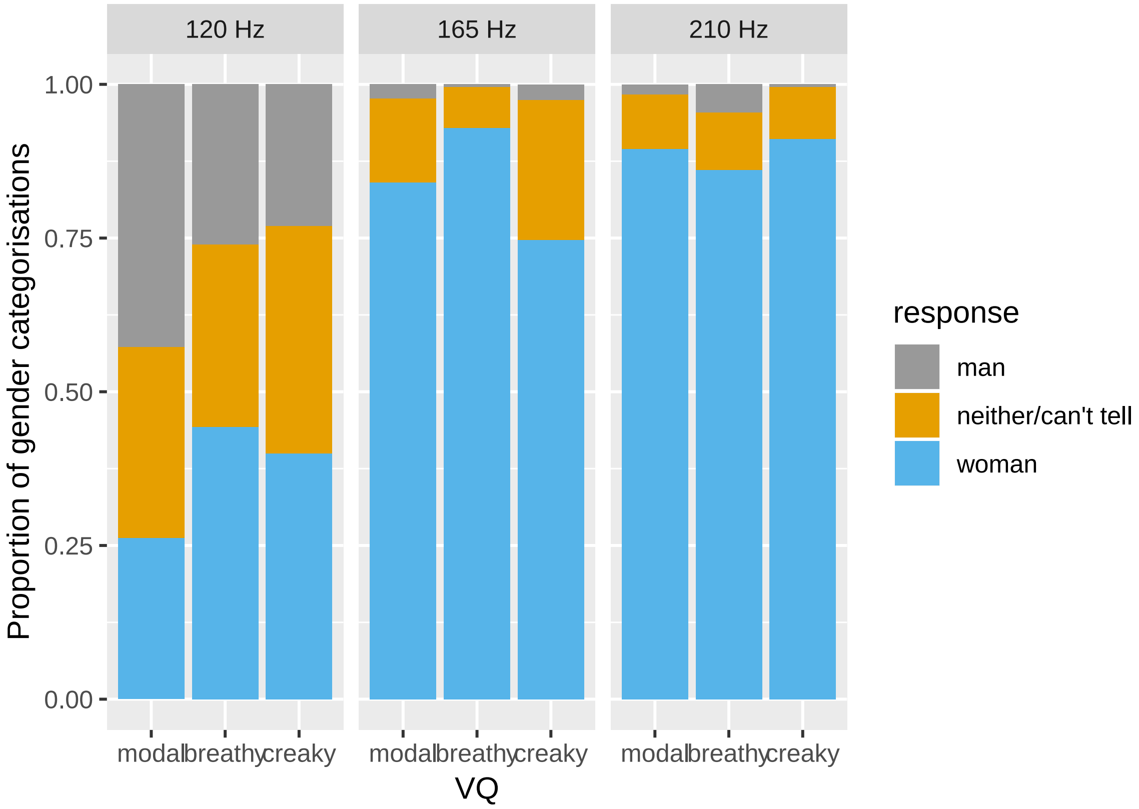


Figure 4.2: Overall distribution of gender responses by voice quality at each fundamental frequency level

categorise a voice as belonging to a man. This could mean that at an ambiguous f0 level, the low f0 of creaky voice has more of an impact on listener categorisation of speaker gender than cultural meaning of creaky voice, which seems, at least in this study, to be 'not a man' and 'not masculine'.

It is interesting to note that although creaky voice at 165Hz led listeners to be less likely to categorize a voice as a woman, it did not lead them to perceive it as any more masculine or any less feminine. In American English, creaky voice has also been identified as a feature of gay-sounding speech (Podesva 2007; Zimman 2013). As many listeners in this study were North American, and even those who were not would have likely had some exposure to the American gay-sounding speech style, perhaps creaky voice at this pitch level indexed a gay male identity for listeners, leading them to be less likely to categorise the voices as women, but not as more masculine or less feminine.

Figure 4.2 shows the proportion of 'man', 'woman', and 'neither/can't tell' responses by voice quality at each f0 level. Examining Figure 4.2 suggests that this finding is most likely driven by high amount of 'neither/can't tell' responses, although there are also a fair number of 'man' responses. One possible interpretation of this is that some listeners are perceiving the ambiguous f0 range combined with a creaky voice as indexing a non-binary identity. In my previous research looking at creaky voice use among transgender speakers (Pearce 2019), one Scottish non-binary speaker used more than twice as much creaky voice as any of the other speakers in the study. While this was not a consistent pattern, with most other speakers using relatively low amounts of creaky voice regardless of identity, there is a possibility that this speaker's use of creaky voice was reflecting a wider pattern among Scottish non-binary speakers that could potentially be being picked up on by listeners, especially given the large number of listeners in the Scottish sample who identified outside of the gender binary.

Breathy voice at 210Hz At first glance, the findings that breathy stimuli were perceived as more masculine and less likely to be perceived as being produced by a woman at 210 Hz, relative to how it was perceived in comparison to modal voice at 120 Hz could indicate that breathy voice is seen as relatively masculine in a voice at a 'typically female' pitch level. However, this seems unlikely in light of previous research on breathy voice that does not indicate that it is perceived as masculine.

One possible interpretation is that at 210Hz, voices are already at a ceiling of being perceived as a woman speaking, and as 'not at all masculine', and so adding in breathy voice does not decrease the perception of masculinity or increase the amount of categorisation as a woman. Furthermore, listeners are likely to be more used to hearing breathy voice in voices at 210 Hz than in lower voices, because of the fact that breathy voice is more common in women's voices. This may then lead to breathy voice being more salient at the 120 Hz and 165 Hz levels, and less salient at the 210 Hz level, leading listeners rating it as more

However, looking at the proportion of 'man', 'woman', and 'neither/can't tell' responses at 210 Hz, as shown in Figure 4.2, suggests this is not whole story: the decrease in 'woman' categorisations at 210 Hz also appears to exist relative to how modal voice is categorised at 210 Hz, not just compared with how breathy voices are categorised in relation to modal voice at 120 Hz.

Potential interactions with KlattGrid synthesis However, these finding could, at least in part, be due to something to do with the synthesis of creaky and breathy voice using KlattGrid. As shown in Figure 4.3, most of the 'not woman' responses that occur in creaky voice at 165 Hz, and in breathy voice at 210 Hz, occur in the stimuli that were synthesised with KlattGrid.

4.1.6 The interaction of listener origin/experiment and fundamental frequency

Summary of predictions and findings I predicted that listeners' origin and domicile might affect how they rate the gender attributes and categorise the gender of the

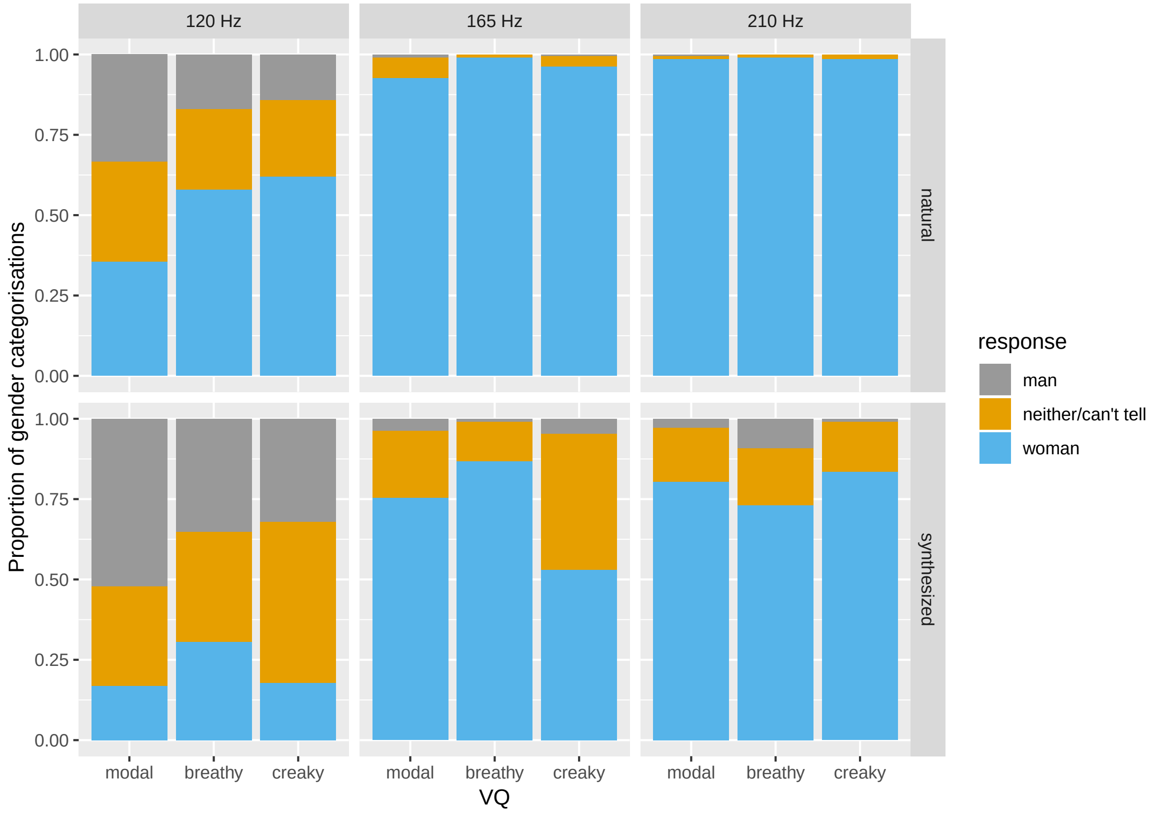


Figure 4.3: Overall distribution of gender responses by voice quality at each fundamental frequency level

speech that they hear, due to culturally-specific gender differences in speech that they are used to hearing in their everyday lives. I did not have any specific expectations about how the influence of listener origin and f0 might interact. Consistent with this, there were differences in how North American listeners rated stimuli in comparison to Scottish listeners at different f0 levels.

Specifically, at 120 Hz (the reference level in the linear mixed-effects model), North American listeners rated stimuli as less masculine, more feminine, and more likely to be a woman than Scottish listeners did. Compared to this, at 165 Hz, North American listeners rated stimuli as more masculine, but showed similar trends in how they categorised speaker gender and rated femininity at 165 Hz as they did at 120 Hz. Again compared to how how North American listeners rated stimuli relative to Scottish listeners at 120 Hz, at 210 Hz, North American listeners rated stimuli as less feminine, more masculine, and were more likely to categorise the speaker as a woman.

120 Hz The biggest difference between Scottish and North American listeners' ratings of gender comes at 120 Hz, so is not readily apparent in the model, due to this being the reference level. However, it is clear that at 120 Hz, North American listeners rate voices and sounding less masculine than Scottish listeners do, and are more likely to categorise the speaker as a woman.

This suggests that there may be some sort of cross-cultural difference in f0 between Scottish and American English. It could be that the typical f0 for a male speaker could be lower in American English than in Scottish English, leading listeners to be less likely to interpret a voice with a median f0 of 120 Hz as a man; on the other hand, it may be that is more usual for female speakers to be speak with a low f0 in American English, leading listeners to be more likely to interpret a voice with a median f0 of 120 Hz as a woman. While I do not know of any research that compares f0 between Scottish English and American English speakers, as we saw in Section 1.1.2, it is common for f0 to vary across cultures and language varieties, and a difference of this kind would explain the difference in how listeners have categorised these voices.

165 Hz & 210 Hz At 165 Hz, North American listeners are still rating stimuli as less masculine than Scottish listeners, but the difference between North American and Scottish listeners has decreased relative to the difference at 120 Hz. However, at this f0 level, there is no significant difference in terms of femininity or gender responses relative to 120 Hz.

At 210 Hz, however, not only does the difference between North American and Scottish listeners decrease even more relative to the difference at 120 Hz, but in terms of gender responses and femininity, they actually reverse, so that North American listeners are less likely than Scottish listeners to categorise the speaker as a woman and rate stimuli as less feminine than Scottish listeners.

Overall, then, Scottish listeners are more extreme in their ratings of gender at 120 Hz and 210 Hz. This suggests that overall, Scottish listeners pay more attention to f0 cues when rating the gender of a speaker, while North American listeners must be paying more attention to other cues.

4.1.7 Discussion of qualitative results

At the end of the experiments, participants were asked the following questions.

1. What did you pay attention to when deciding whether you thought the voices sounded like a man or a woman?

2. What did you pay attention to when rating how masculine/feminine a voice sounded?

3. Do you have any further comments about your experience completing this experiment?

Participants' answers to these questions suggested a degree of awareness of the topic of the experiment and of voice quality as a cue to speaker gender. While pitch was the most common cue that participants mentioned, 15 responses to Question 1 and 20 responses to Question 2 used language that suggested they had paid attention to voice quality. This difference in how often it was mentioned in Question 2 versus Question 1 suggests that participants were more aware of paying attention to voice quality cues when rating how masculine or feminine a voice sounded, suggesting they might see voice quality as more connected to the gender attributes of a speaker than to speaker gender itself.

The responses of some participants suggest that they often rated gender in conjunction with other aspects of speaker identity, taking into account whether they thought the speaker was gay, a smoker, how old they were and their accent. This suggests that gender perception for these participants was closely related to their perception of other traits in the voice. As these speaker characteristics are also things that voice quality can index, this shows that it is important to take the other identities that voice quality may index into account when investigating how voice quality relates to gender perception.

4.2 Overall discussion

4.2.1 Main research questions

How does voice quality affect perceptions of speaker gender?

Previous research on the contribution of voice quality had generally found that breathy voice was perceived as feminine, while creaky voice was perceived as masculine, but these findings were not conclusive.

In this study, I conducted a perception experiment using stimuli with naturally produced and synthesized voice quality differences, where listeners were asked to rate the masculinity and femininity of voices and categorised the speaker as either a 'man', 'woman' or as 'neither/can't tell'.

I found that relative to modal voice, listeners were more likely to categorise breathy voice stimuli as women, and rated breathy voice stimuli as more feminine and less masculine. This finding is broadly consistent with previous research. The finding that breathy voices are not only perceived as more feminine, but also more likely to be heard as women's voices, has potential implications for feminizing vocal therapy, as depending on the needs of the client, the goal of this therapy is often not only to develop a more feminine-sounding voice, but also to develop a voice that is perceived as belonging to a woman by listeners.

On other hand, this finding also has potential implications for masculinizing vocal therapy, which is attributed less attention due to the effects of testosterone on the voice, but is often sought in cases where individuals may not want or be able to undergo hormone replacement therapy, or where they have started hormone replacement therapy and have not experienced the desired vocal changes (Davies et al. 2015). Achieving a more masculine or male-sounding voice in vocal therapy may involve aiming to lower pitch and vowel formant frequencies (Davies et al. 2015). One potential method of achieving these goals would be through speaking with a lowered larynx, which increases the size of the resonating cavity above the larynx to lower formant frequencies, and is generally accompanied by a lowering of fundamental frequency (Laver 1980: 301). However, the muscles used to lower the larynx interact with those used to produce breathy voice, meaning that speaking with a lowered larynx also frequently means speaking with a breathy voice quality (Laver 1980: 31). The finding that listeners are more likely to categorise a breathy voice as a woman, and rate breathy voice as more feminine and less masculine, should thus be taken into account when considering this approach in masculinizing vocal therapy.

However, I found that although they did not rate creaky voice as more feminine than modal voice, listeners rated creaky voice stimuli less masculine and were more likely to categorise creaky voice stimuli as women relative to modal voice. This is interesting in light of the fact that previous research conducted in Scotland (Beck and Schaeffler 2015; Stuart-Smith 1999) and other parts of the UK more generally (Gittelson et al. 2018; Henton and Bladon 1988) has found that male speakers are generally more creaky. However, work conducted in London has found that creaky voice was more common among some groups of female speakers than their male counterparts (Szakay and Torgersen 2015, 2019). The findings of this study, then, potentially suggest that gender differences in voice quality production may be undergoing change, with female speakers in Scotland becoming more creaky, as has been a trend in American English speakers and among certain groups in London.

Podesva (2013) suggested that there may be an iconic link between creaky voice and masculinity due to the low f0 of creaky voice being associated with the low f0 typically used by male speakers. This was, to an extent, supported by previous research by Greer (2015) and Lee (2016) who found that despite increasing prevalence of creaky voice among young female speaker of American English, listeners still perceived creaky voice as masculine, although Greer (2015) stressed that other results from her study indicated that the link between creaky voice and masculinity was not an iconic one. While it may be the case that an iconic link exists theoretically, the result here that listeners rate creaky voice as less masculine than modal voice suggests that the iconic link between creaky voice and masculinity has somehow been reworked at higher orders of indexicality (Silverstein 2003), and has become associated with women and a lack of masculinity.

At the end of the experiment, listeners were asked what they paid attention to when deciding if they thought a voice belonged to a man or a woman, and when rating how masculine or feminine a voice sounded. A lot of listeners used words that suggested they were paying attention to voice quality when rating voices - in Question 1, 15 participants used language that suggested they had paid attention to voice quality, while in Question 2, 20 participants used language that suggested they had paid attention to voice quality. One speaker specifically noted that a breathy voice made the voice sound feminine, saying that, 'some of the speakers let out slightly more breath as they were finishing their last word. This made it seem much more feminine'. Additionally, many listeners noted that they thought some voices had a 'soft' quality that made them sound more feminine or more like women than men - while this could be a reference to loudness, given the quantitative evidence suggesting that listeners paid attention to breathy voice, it seems like listeners could be using 'soft' to mean breathy. Furthermore, listeners also referenced that they had paid attention to 'vocal fry', 'breathy-ness', 'the warmth of the voice', 'how much rich the sound was coming from the larynx', 'harshness', 'nasality', 'raspiness', and how 'gruff' or 'hard' it sounded.

As breathy and creaky voice were both less likely to be categorised as women and rated less masculine than modal voice, it seems that in this study, modal voice has emerged as a masculine trait that can suggest a man is talking. One participant echoed this in their response to Question 1, noting that 'men's voices frequently sound a bit sharper than women's'.

However, the results of this study also show that voice quality is secondary to other cues when it comes to gender perception. Residual cues to gender in this speaker's voice, presumably mostly vowel formant frequencies, were very important to listeners perception of gender. In Question 3, 3 participants commented that they had categorised very few voices as men, or that the study should have used more men's voices, suggesting that despite attempts to alter gender perception through manipulation of f0 and voice quality, residual cues to gender still played an important role. This is supported by the fact that there was a strong overall bias towards listeners categorising voices as 'women'. F0 level also had a large effect on gender perception, which will be discussed in more detail in the next section.

Overall, the results of this study suggest that breathy voice and creaky voice are both voice qualities that cue listeners to perceive voices as non-masculine and as belonging to a woman. These finding have implications for vocal therapy for transgender clients, give evidence that a shift in gender patterns in voice quality production may have occurred in Scotland, and suggest that for listeners, the iconic link between masculinity and creaky voice due to low f0 has been lost.

How does voice quality interact with fundamental frequency when cuing gender perceptions?

Previous research has not only found gender differences in voice quality in terms of production, but has also found that certain voice qualities are perceived differently in the voices of male and female speakers (Greer 2015) and that voice quality is a more important cue to speaker gender where other cues, such as f0, are ambiguous (Skuk and Schweinberger 2014). Manipulating the median f0 of the stimuli to either a 'typical male' (120Hz), 'typical female' (210Hz) or 'androgynous' level (165 Hz) allowed investigation of how perceptions of voice quality interacted with pitch, an important cue to speaker gender in the voice.

In itself, f0 was a strong cue to gender perception, with higher f0s levels leading listeners to be more likely to categorise voices as women. However, a f0 level alone was not the only cue to speaker gender, and at the 120 Hz level, intended to reflect a 'typical male' pitch, just 31% of stimuli were categorised as being a 'man', while 33% were categorised as 'neither/can't tell', and 37% were categorised as being a 'woman'. This suggests that although f0 is an important cue to speaker gender, residual cues, likely in the form of vowel formant frequencies, are also particularly important.

While in Section 4.1.5 we already discussed how voice quality interacted with pitch from a quantitative perspective, insight from the qualitative data also suggests that listeners took pitch into account when evaluating voice quality.

Firstly, one person noted that they rated voices as 'masculine' because they could imagine gay men they knew 'saying it in that tone/pitch/whatever it's called'. This is particularly interesting in light of how listeners categorised creaky voices at 165Hz, as more likely to be a man, but not different in terms of masculinity or femininity, suggesting that creaky voice at this ambiguous pitch level may be interpreted by listeners as indexing a gay male identity, as previous research has found it to be a feature of gay-sounding speech.

Additionally, several speakers noted that they paid attention to what they thought the speaker's age was or whether the speaker was a smoker when rating their gender. For example, one participant noted that 'What might have been young female could well have been late adolescent male. Also, it was challenging to guess between possible post-menopausal female versus male with high-registered voices.', while another said that 'Some of the lower-pitch, "raspy" voices sounded more male, but only slightly so like a female smoker'. As age and smoking are both things which can affect voice quality and pitch of the voice, these comments suggest that listeners considered voice quality and pitch together to rate or categorise the gender of a speaker.

Overall, the qualitative data collected suggests that participants not only took pitch into account when rating the gender, but that the pitch of the voice affected how they interpreted other cues to gender such as voice quality in relation to speaker gender.

4.2.2 Secondary research questions

Does listener's sociolinguistic background and exposure to culturally-specific gender differences in voice quality affect whether voice quality contributes to gender perception?

Some previous work had found that listeners with different sociolinguistic backgrounds affects the ways that voice quality contributes to gender perception (Foulkes et al. 2011). However, most previous work looking at voice quality perception did not take the varieties spoken by their speakers and backgrounds of their listeners into account.

In the present study, the interaction between voice quality and listener origin was not included in any of the final models after the full model was stepped down to only include interactions that improved the model, suggesting that the effects of voice quality are consistent between Scottish and North American listeners. This is contrary to predictions, as I expected that cross-cultural gender differences in production might influence how voice quality was perceived.

However, this is perhaps unsurprising, as the listeners from Experiment 1 and Experiment 2 were all English speakers, and the gender differences in production are not that extreme — female speakers use more breathy voice in both North American and UK varieties, and work on varieties of English in London has found that some groups female speakers are using more creaky voice too (Szakay and Torgersen 2015, 2019). Therefore, in future research it would be interesting to examine the contribution of voice quality to speaker gender between varieties of languages where the cross-cultural difference in voice quality is more robust.

In addition to this, further attention should be paid to how different aspects of sociolinguistic background might influence listeners' perception of gender. Following completion of the experiment, several of the local participants gave additional comments on the experiment and two of them asked if there were transgender speakers among the voices that they had heard. One had worked as a speech and language therapy assistant with transgender clients and said that one of the voices sounded like a transgender woman that she had worked with, while another said that she thought

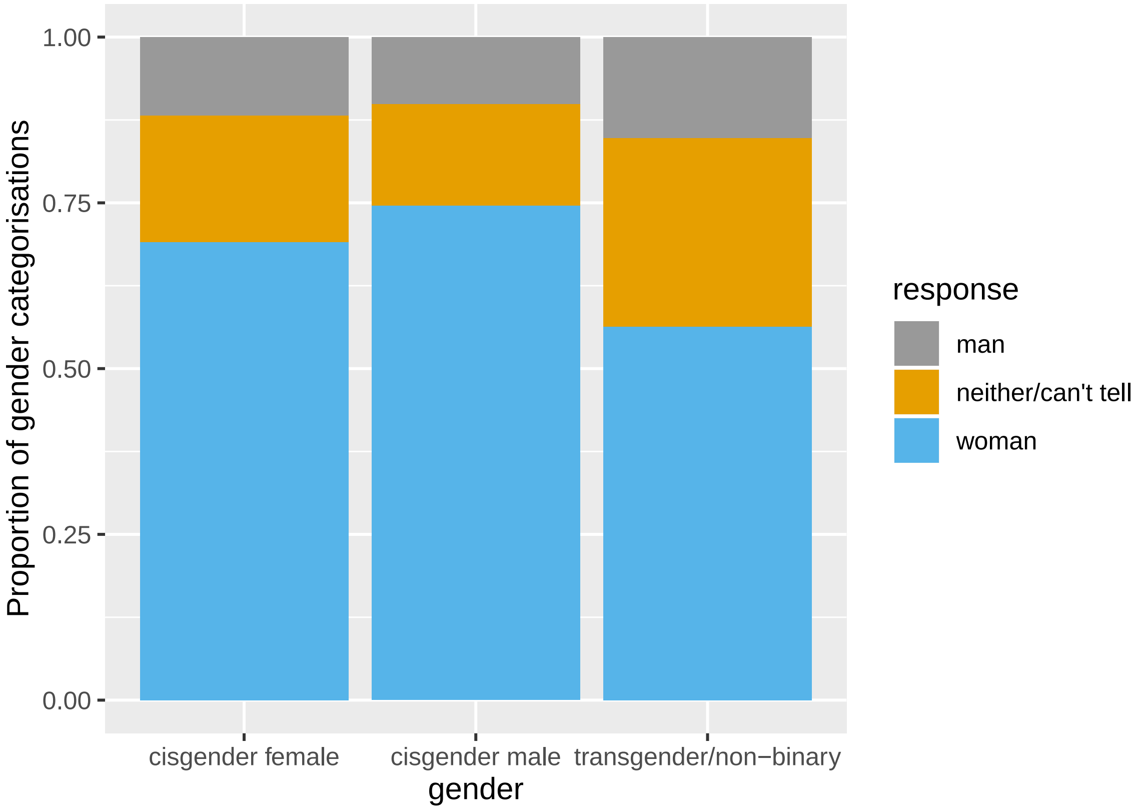


Figure 4.4: Overall distribution of gender responses by listener gender

that one of the voices reminded her of transgender men that she knew – and that they did not sound like cisgender men she knew from Glasgow because they did not speak in the same 'tough' way [1]. How much exposure a participant has had to the voices of transgender and gender non-conforming individuals may well influence the way that they think about gender in the voice, and thus how they categorise gender and rate gender attributes in voices.

Going into this study, I had no expectations about how a listener's own gender might influence their perception of voices and therefore did not include listener gender in my statistical models. However, observing the distribution of listener's categorisations of speaker gender by listener gender reveals that listener gender may have some kind of influence. Figure 4.4 shows categorisations of speaker gender by listener gender, which has been collapsed from the more detailed information about listener gender that was collected into three categories, 'cisgender female', 'cisgender male', and 'transgender/non-binary', which includes participants with a range of non-binary identities, one transgender woman and one transgender man. Figure 4.5 shows categorisations of speaker gender by listener gender, broken down by the f0 level of the stimuli.

First of all, there seems to be a slight trend overall where cisgender men categorise slightly more stimuli as women, and slightly less as men, relative to the responses of

[1] Both of these participants gave their consent to these additional comments being included here

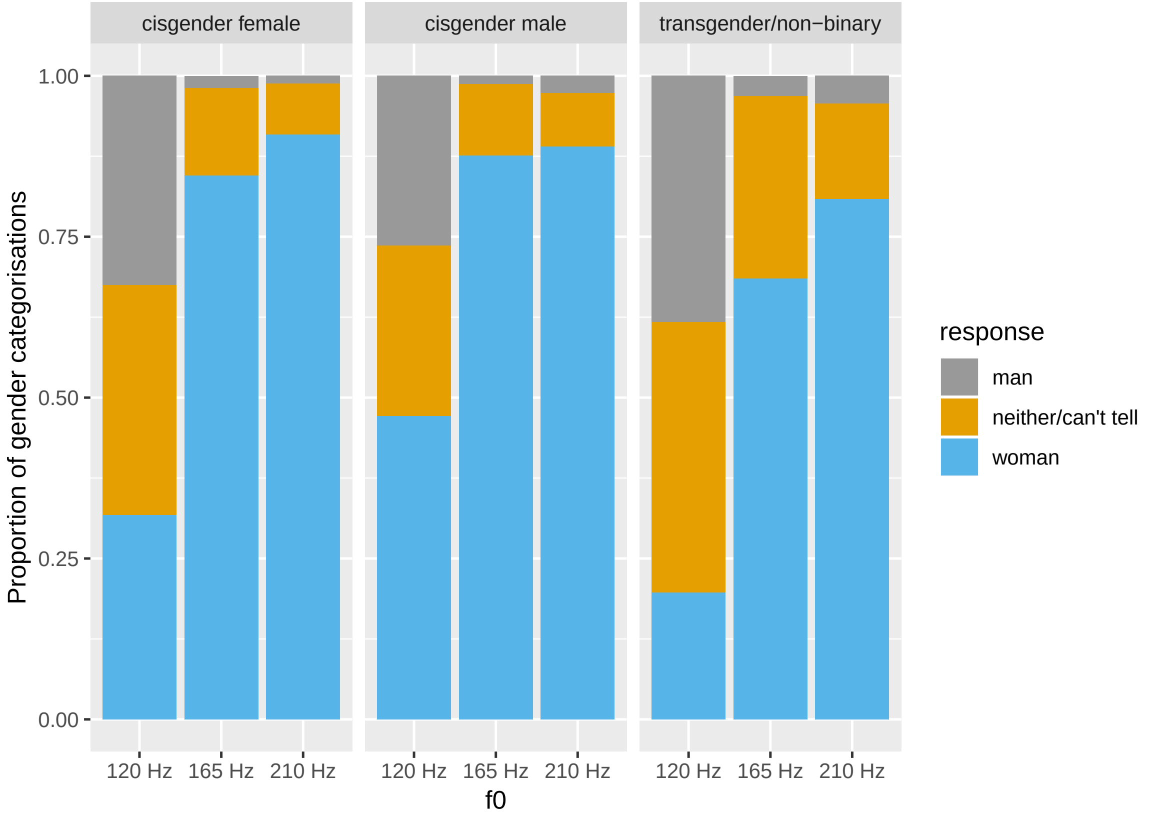


Figure 4.5: Overall distribution of gender responses by listener gender at each f0 level

cisgender women. At 120Hz, this trend is more apparent. This could indicate that cisgender men are less willing than cisgender women to categorise voices as being men when they lack characteristics that they except to hear in a male voice, despite cues of f0. It is worth noting that there may also be a confound between listener gender and listener origin/experiment. Experiment 1 involved 33 participants: 9 cisgender male, 16 cisgender female and one transgender female participant, as well as 7 who selected 'other' and wrote in a different identity. Experiment 2 involved 40 participants: 17 cisgender female, 22 cisgender male, and 1 transgender male participant. It is thus unclear whether these differences in responses between the cisgender male and cisgender female participants reflect a difference to do with their gender, their origin, or the conditions in which they completed the experiment.

Furthermore, compared to the two groups of cisgender speakers, transgender/non-binary listeners seem to categorise less stimuli as women, but more as men and as 'neither/can't tell'. That transgender/non-binary listeners categorise more stimuli as men could indicate that through an awareness of what transgender speakers sound like, listeners of this group are more willing to accept voices as men, even if they display residual cues to gender that might suggest the speaker is not male. This trend is more apparent at 120 Hz. Furthermore, the fact that transgender/non-binary listeners categorise more stimuli as 'neither/can't tell' may reflect a wider unwillingness to assume the gender of a speaker based on their voice, something which is reflected in the feedback given by one agender participant in Question 3, ‘I didn’t know how to rate things in terms of masculinity and femininity, or whether it was a man or a woman, because I don’t feel like I can know someone’s gender by hearing their voice’.

Overall, while sociolinguistic background, in the form of listener origin, did not affect how voice quality influences perception of gender, listener origin did affect the role of fundamental frequency in gender perception. Furthermore, there is tentative evidence that other aspects of sociolinguistic background may play a role in gender perception, and this should be investigated further in future research.

How does the type of response used to measure ‘perceived gender’ affect whether voice quality contributes to gender perception?

Leung et al. (2018) notes that some of the variation between studies in what contributes to gender perception can be explained in terms of the type of response that researchers used to measure ‘perceived gender’. I speculated that perhaps voice quality might have influence perception of a voice as masculine or feminine, but not on perception of speaker gender identity. I found that voice quality had an effect on each of these different ways of measuring ‘perceived gender’, but that there were differences in the exact effect that it had.

The fact that relative to modal voice, listeners were more likely to categorise creaky voice was spoken by a woman, and perceived as less masculine, but *not* more feminine, suggests that these scales elicit different responses from participants. Furthermore, North American listeners were more likely to categorise stimuli as spoken by a woman and rated stimuli as less masculine overall than Scottish listeners did. However, there was no significant difference in how feminine Scottish and North American listeners rated results. This suggests that not only do listener categorisation of gender and scales of masculinity and femininity function independently of each other, but that perhaps there are differences in what these terms mean to different listeners and that this may vary between listeners of different backgrounds.

On the one hand, some evidence from the feedback questions suggests that some listeners saw the different kinds of responses as equivalent. In response to Question 2, one participant said ‘I paid attention to the same things when deciding whether man or woman, masculine or feminine’, another said ‘same as before’, and another said ‘higher rating came when i was more sure of the gender ’. Furthermore, in response to Question 3, one participant said ‘i would have rather the study stuck to just using the meter of how masculine/feminine the voices sounded rather than asking if it was a man or woman or neither.

On the other hand, certain participants noted that to them, masculinity, femininity and the gender of the speaker were independent of each other. For example, in response to Question 3, one participant noted ‘Sometimes a more masculine voice meant a less feminine one but not always so it was good to have the two sliders’. Furthermore, another participant noted that they would have liked to be able to express more nuance when categorising the gender of a speaker: ‘there should have been an option for “women speaker but with a masculine voice”, or some more indication that this could be a possibility’.

Furthermore, some participants noted than the traits they had been asked to had not been specified well. For example, in response to Question 3, one participant said ‘it was not clear if feminine/masculine meant “female/male” or socially masculine/feminine’.

Overall, it seems that participants do not see femininity, masculinity, and gender identity to be equivalent, and the terms that I used to measure perceived gender mean different things to different listeners.

4.3 Discussion of methodology

Synthesis The KlattGrid synthesis that was used to create synthesized differences in voice quality may have led to some confounds in the results. Firstly, in some cases it was difficult to tell whether results stemmed from actual differences in how the voice quality was perceived, or differences in how natural the synthesised stimuli in a certain condition sounded.

Furthermore, although I sought the opinions of my supervisors and other people in my department as to whether I had successfully synthesised the desired voice qualities, I was not able to conduct a full norming study to see whether the KlattGrid settings that I had used were successful in creating the target voice qualities.

Additionally, I was unable to find a male speaker whose voice worked well with the synthesis. In future research, it would be beneficial to include a male speaker as well to counterbalance the stimuli. I would recommend using a speaker who can successfully produce modal voice to help this. Additionally, it would likely be beneficial to follow advice given in the Praat manual for creating a successful formant object, asking the speaker to produce steady state vowels and using a maximum formant value that works for the individual speaker Boersma and Weenink (2019). I suspect that my attempts to use the default value of 5000Hz for the maximum formant for my male speakers might have influenced the output of the KlattGrid, and that following this advice might have yielded a better result.

Additionally, KlattGrid copy synthesis does come at the cost of naturalness in the stimuli. Future research might consider investigating similar questions using Gobl and Ní Chasaide (2003)'s method of manually inverse filtering an utterance to produce more natural sounding synthesis of voice quality, or alternatively using the UCLA voice synthesiser 2.0 (Kreiman and Gerratt 2005) to investigate the role of voice quality in gender perception in steady state vowels.

Natural stimuli On the other hand, the synthesised stimuli were more controlled than the natural stimuli. The natural stimuli may also have caused confounds in the research; while the speaker did succeed in producing three distinct voice qualities, it may be the case that in doing so she also altered other aspects of her voice, such as prosody, that were not controlled for.

Beyond creaky and breathy voice Production research has found important gender differences not just in breathy and creaky voice, but in a wide range of phonation types including whispery voice, laryngealised/tense/pressed voice, and falsetto. Future research should investigate the contribution of each of these aspects of voice quality, and more, to perception of gender, especially where production research has found locally-situated gender differences in these phonation types.

Chapter 5

Conclusion

In this study, I found that voice quality influences listener perception of speaker gender. Relative to their ratings of modal voice, listeners were more likely to categorise breathy voice and creaky voice stimuli as women, rated breathy voice and creaky voice as less masculine, and rated breathy, though not creaky voice, as more feminine. These findings suggest that modal voice stands out as a masculine-sounding voice quality that leads listeners be less likely to think the voice belongs to a woman.

These findings also suggest that the gendered meanings of creaky voice may be changing in Scottish English. Creaky voice was perceived as less masculine than modal voice, suggesting that it is no longer linked to masculinity through its low f0. This could reflect a wider trend for creaky voice to be more prevalent in female speakers, which has been found in London and North America, but never in speakers of Scottish English.

Through considering qualitative data, listener background, the different types of response used to measure gender perception, and the interaction between fundamental frequency and voice quality perception, I also suggest that listener's perception of gender is mediated by their experiences and background. Future research should bear this in mind when investigating the contribution of voice quality to the perception of speaker gender, and take into account locally-situated differences in the gendered meanings of voice quality that have been found in previous research looking at voice quality production.

Bibliography

Abdelli-Beruh, Nassima B., Lesley Wolk, and Dianne Slavin (2014). “Prevalence of vocal fry in young adult male american english speakers”. In: *Journal of Voice* 28 (2), pp. 185–190.

Abercrombie, David (1967). *Element of General Phonetics.* Edinburgh: Edinburgh University Press.

Addington, David W. (1968). “The relationship of selected vocal characteristics to personality perception”. In: *Speech Monographs* 35 (4), pp. 492–503.

Andrews, Moya L. and Charles P. Schmidt (1997). “Gender presentation: Perceptual and acoustical analyses of voice”. In: *Journal of Voice* 11 (3), pp. 307–313.

Avery, Jack D. and Julie M. Liss (1996). “Acoustic characteristics of less-masculine-sounding male speech”. In: *The Journal of the Acoustical Society of America* 99 (6), pp. 3738–3748.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). “Fitting Linear Mixed-Effects Models Using lme4”. In: *Journal of Statistical Software* 67 (1), pp. 1–48.

Beck, Janet (1988). “Organic variation and voice quality”. PhD thesis. University of Edinburgh.

Beck, Janet (2005). “Perceptual analysis of voice quality: The place of the Vocal Profile Analysis”. In: *A Figure of Speech: A festschrift for John Laver.* Ed. by Janet Beck and William J. Hardcastle. Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 285–232.

Beck, Janet and Felix Schaeffler (2015). “Voice quality variation in Scottish adolescents : gender versus geography”. In: *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015), Glasgow, Scotland.*

Becker, Kara, Sameer Ud Dowla Khan, and Lal Zimman (2017). “Creaky voice beyond binary gender”. In: *American Dialect Society Annual Meeting 2017, Austin, Texas.*

Bishop, Jason and Patricia Keating (2012). “Perception of pitch location within a speaker’s range: Fundamental frequency, voice quality and speaker sex”. In: *The Journal of the Acoustical Society of America* 132 (2), pp. 1100–1112.

Boersma, Paul (1993). “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound”. In: *Proceedings of the Institute of Phonetic Sciences.* University of Amsterdam, pp. 97–110.

Boersma, Paul and David Weenink (2019). *Praat: doing phonetics by computer.* [Computer program]. Version 6.0.49, Amsterdam.

Booz, Jaime A. and Sarah H. Ferguson (2016). "Perceived gender in clear and conversational speech". In: *The Journal of the Acoustical Society of America.*

Bradley, Harriet (2014). *Gender*. 2nd edition. Cambridge: Polity Press.

Bucholtz, Mary (2002). "From 'sex differences' to gender variation in sociolinguistics". In: *U. Penn Working Papers in Linguistics* 8 (3).

Busby, P. A. and G. L. Plant (1995). "Formant frequency values of vowels produced by preadolescent boys and girls". In: *The Journal of the Acoustical Society of America* 97 (4), pp. 2603–2606.

Butler, Judith (2014[1990]). *Gender trouble: Feminism and the subversion of identity.* New York / London: Routledge.

Butler, Judith (2014[1993]). *Bodies that matter: On the discursive limits of 'sex'.* New York / London: Routledge.

Cartei, Valentina, Rod Bond, and David Reby (2014). "What makes a voice masculine: Physiological and acoustical correlates of women's ratings of men's vocal masculinity". In: *Hormones and Behavior* 66 (4), pp. 569–576.

Catford, J. C. (1977). *Fundamental problems in phonetics.* Edinburgh: Edinburgh University Press.

Coleman, Ralph O. (1976). "A Comparison of the Contributions of Two Voice Quality Characteristics to the Perception of Maleness and Femaleness in the Voice". In: *Journal of Speech and Hearing Research* 19 (1), pp. 168–180.

Dallaston, Katherine and Gerard Docherty (2019). "Estimating the prevalence of creaky voice: a fundamental frequency-based approach". In: *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019.* Ed. by Marija Tabain Sasha Calhoun Paola Escudero and Paul Warren. Canberra, Australia: Australasian Speech Science and Technology Association Inc., pp. 532–536.

Davies, Shelagh and Joshua Goldberg (2006). *Transgender Speech Feminization/Masculinization Suggested Guidelines for BC Clinicians.* Vancouver: Vancouver Coastal Health, Transcend Transgender Support & Education Society, and the Canadian Rainbow Health Coalition.

Davies, Shelagh, Viktória G. Papp, and Christella Antoni (2015). "Voice and Communication Change for Gender Nonconforming Individuals: Giving Voice to the Person Inside". In: *International Journal of Transgenderism* 16 (3), pp. 117–159.

Di Paolo, Marianna and Alice Faber (1990). "Phonation differences and the phonetic content of the tense-lax contrast in Utah English". In: *Language Variation and Change* 2 (2), pp. 155–204.

Dominic Mazzoni (2018). *Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.3.2.* Audacity Team.

Dorreen, Keiran (2016). "Fundamental frequency distributions of bilingual speakers in forensic speaker comparison". MA thesis. University of Canterbury.

Eckert, Penelope (2000). *Linguistic variation as social practice: the linguistic construction of identity in Belten High*. Malden, MA: Blackwell.

Eckert, Penelope (2001). "Style and social meaning". In: *Style and sociolinguistic variation*. Ed. by Penelope Eckert and John R. Rickford. New York: Cambridge University Press, pp. 119–126.

Eckert, Penelope (2008). "Variation and the indexical field". In: *Journal of Sociolinguistics* 12 (4), pp. 453–476.

Eckert, Penelope (2017). "Comment: The most perfect of signs: Iconicity in variation". In: *Linguistics* 55 (5), pp. 1197–1207.

Eckert, Penelope and Sally McConnell-Ginet (2013). *Language and gender*. English. Second. Cambridge, UK: Cambridge University Press.

Esling, John (1978). "The identification of features of voice quality in social groups". In: *Journal of the International Phonetic Association* 8 (1-2), pp. 18–23.

Esling, John and Jimmy Harris (2005). "States of the Glottis: An articulatory phonetic model based on laryngoscopic observations". In: *A Figure of Speech: A festschrift for John Laver*. Ed. by Janet Beck and William J. Hardcastle. Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 347–384.

Fant, Gunnar (1960). *Acoustic theory of speech production: with calculations based on X-ray studies of Russian articulations*. English. Vol. 2. s-Gravenhage: Mouton.

Ferrand, Carole T and Ronald L. Bloom (1996). "Gender Differences in Children's intonational patterns". In: *Jounral of Voice* 10 (3), pp. 284–291.

Foulkes, Paul, Gerry Docherty, Ghada Khattab, and Malcah Yaeger-Dror (2011). "Sound judgements: perception of indexical features in children's speech". English. In: *A reader in sociophonetics*. Ed. by Dennis R. Preston and Nancy A. Niedzielski. New York, N.Y: De Gruyter Mouton.

Fox, John and Sanford Weisberg (2019). *An R Companion to Applied Regression*. 3rd. Thousand Oaks CA: Sage.

Gallena, Sally J. K., Betsy. Stickels, and Emily. Stickels (2018). "Gender Perception After Raising Vowel Fundamental and Formant Frequencies: Considerations for Oral Resonance Research". In: *Journal of Voice* 32 (5), pp. 592–601.

Garellek, Marc (2019). "The phonetics of voice". In: *The Routeledge Handbook of Phonetics*. Ed. by William F. Katz and Peter F. Assmann. Abingdon, Oxon;New York, NY; Routledge.

Gelfer, Marylou Pausewang and Quinn E. Bennett (2013). "Speaking fundamental frequency and vowel formant frequencies: Effects on perception of gender". In: *Journal of Voice* 27 (5), pp. 556–566.

Gelfer, Marylou Pausewang and Victoria A. Mikos (2005). "The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels". In: *Journal of Voice* 19 (4), pp. 544–554.

Gelfer, Marylou Pausewang and Kevin J. Schofield (2000). "Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male". In: *Journal of Voice* 14.1, pp. 22–33.

Gerratt, Bruce and Jody Kreiman (2001). "Measuring vocal quality with speech synthesis". In: *The Journal of the Acoustical Society of America* 110 (5), pp. 2560–2566.

Gittelson, Ben, Yang Li, and Adrian Leemann (2018). "Acoustic analysis of nonmodal phonation in 1000+ speakers from across the UK". In: *The 2018 Colloquium of the British Association of Academic Phoneticians.*

Gobl, Christer and Ailbhe Ní Chasaide (2003). "The role of voice quality in communicating emotion, mood and attitude". In: *Speech Communication* 40 (1-2), pp. 189–212.

Gobl, Christer and Ailbhe Ní Chasaide (2010). "Voice source variation and its communicative functions". In: *The Handbook of Phonetic Sciences.* Ed. by William J. Hardcastle, John Laver, and Fiona E. Gibbon. 2nd ed. Oxford: Blackwell, pp. 378–423.

Gobl, Christer, Irena Yanushevskaya, Andy Murphy, and Ailbhe Ní Chasaide (2019). "Comparison of time and frequency domain measures of the voice source". In: *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019.* Ed. by Marija Tabain Paul Warren Sasha Calhoun Paola Escudero. Canberra, Australia: Australasian Speech Science and Technology Association Inc., pp. 3290–3294.

Gordon, Matthew and Peter Ladefoged (2001). "Phonation types: a cross-linguistic overview". In: *Journal of Phonetics* 29 (4), pp. 383–406.

Gorham-Rowan, Mary and Richard Morris (2006). "Aerodynamic Analysis of Male-to-Female Transgender Voice". In: *Journal of Voice* 20 (2), pp. 251–262.

Gray, Henry (1918). *Gray's Anatomy.* Ed. by Warren H. Lewix. 20th ed. Philadelphia and New York: Lea and Febiger.

Greer, Sarah Doris Faye (2015). "The Perception of Coolness: Voice Quality and Its Social Uses and Interpretations". MA thesis. University of Calgary.

Hancock, Adrienne, Lindsey Colton, and Fiacre Douglas (2014). "Intonation and gender perception: Applications for transgender speakers". In: *Journal of Voice* 28 (2), pp. 203–209.

Hancock, Adrienne, Julianne Krissinger, and Kelly Owen (2011). "Voice Perceptions and Quality of Life of Transgender People". In: *Journal of Voice* 25 (5), pp. 553–558.

Hansom, H. M. (1997). "Glottal characteristics of female speakers: Acoustic correlates". In: *Journal of Acoustical Society of America* 101 (1), pp. 466–481.

Hansom, H. M. and E.S Chuang (1999). "Glottal characteristics of male speakers: Acoustic correlates and comparisons with female data". In: *Journal of Acoustical Society of America* 106 (2), pp. 1064–1077.

Hardy, Teresa L.D., Carol A. Boliek, Kristopher Wells, Carol Dearden, Connie Zalmanowitz, and Jana M. Rieger (2016). “Pretreatment acoustic predictors of gender, femininity, and naturalness ratings in individuals with male-to-female gender identity”. In: *American Journal of Speech-Language Pathology* 25 (2), pp. 125–137.

Hasek, Carol S., Sadanand Singh, and Thomas Murry (1980). “Acoustic attributes of preadolescent voices”. In: *The Journal of the Acoustical Society of America* 68 (5), pp. 1262–1265.

Henton, Caroline (1995). “Cross-language variation in the vowels of female and male speakers”. In: *Proc. XIIIth ICPhS*. Vol. 4. Stockholm, pp. 420–423.

Henton, Caroline and Anthony Bladon (1985). “Breathiness in normal female speech: Inefficiency versus desirability”. In: *Language and Communication* 5 (3), pp. 221–227.

Henton, Caroline and Anthony Bladon (1988). “Creak as a sociophonetic marker”. In: *Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin*. Ed. by L.M. Hyman and C.N. Li, pp. 3–29.

Hillenbrand, James (2011). “Acoustic Analysis of Voice: A Tutorial”. In: *Perspectives on Speech Science and Orofacial Disorders* 21 (2), pp. 31–43.

Hillenbrand, James and Michael Clark (2009). “The role of f0 and formant frequencies in distinguishing the voices of men and women”. In: *Attention, Perception, & Psychophysics* 71 (5), pp. 1150–1166.

Hillenbrand, James, R.A. Cleveland, and R.L. Erickson (1994). “Acoustic correlates of breathy vocal quality”. In: *Journal of Speech and Hearing Research* 37, pp. 769–778.

Hinton, Leanne, Johanna Nichols, and John J. Ohala (1994). “The frequency code underlies sound-symbolic use of voice pitch”. English. In: *Sound symbolism*. Cambridge [England];New York, NY; Cambridge University Press, pp. 325–347.

Hollien, Harry (1960). “Vocal pitch variation related to changes in vocal fold length”. In: *Journal of Speech, Language and Hearing Research* 3 (2), pp. 150–156.

Hollien, Harry, Rachel Green, and Karen Massey (1994). “Longitudinal research on adolescent voice change in males”. In: *The Journal of the Acoustical Society of America* 96 (5), pp. 2646–2654.

Holmberg, Eva B., Robert E. Hillman, Joseph S. Perkell, Peter C. Guiod, and Susan L. Goldman (1995). “Comparisons Among Aerodynamic, Electroglottographic, and Acoustic Spectral Measures of Female Voice”. In: *Journal of Speech, Language, and Hearing Research* 38 (6), pp. 1212–1223.

Holmberg, Eva B., Jennifer Oates, Georgia Dacakis, and Cameron Grant (2010). “Phonetograms, aerodynamic measurements, self-evaluations, and auditory perceptual ratings of male-to-female transsexual voice”. In: *Journal of Voice* 24 (5), pp. 511–522.

Hoofring, Alan (2003). *Larynx and Nearby Structures*. URL: `https://en.wikipedia.org/wiki/File:Larynx_and_nearby_structures.jpg`.

Hubbard, Daniel J. and Peter F. Assmann (2013). "Perceptual adaptation to gender and expressive properties in speech: The role of fundamental frequency". In: *The Journal of the Acoustical Society of America.*

Ingrisano, Dennis, Gary Weismer, and Gordon H. Schuckers (1980). "Sex identification of preschool children's voices". In: *Folia Phoniat* 32, pp. 61–69.

Jackson, Stevi and Sue Scott (2001). "Putting the body's feet on the ground: Towards a sociological reconceptualization of gendered and sexual embodiment". In: *Constructing gendered bodies.* Ed. by Kathryn Backett-Milburn and Linda McKie. Basingstoke: Palgrave, pp. 7–24.

Kawahara, Hideki and Masanori Morise (Oct. 2011). "Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework". In: *Sadhana* 36 (5), pp. 713–727.

Keating, Patricia, Marc Garellek, and Jody Kreiman (2015). "Acoustic properties of different kinds of creaky voice". In: *Proceedings of the International Congress of Phonetic Science XVIII.*

Kiesling, Scott F. (2019). *Language, Gender and Sexuality.* New York: Routledge.

King, Robert S., George R. Brown, and Christopher R. McCrea (2012). "Voice parameters that result in identification or misidentification of biological gender in male-to-female transgender veterans". In: *International Journal of Transgenderism* 13 (3), pp. 117–130.

Klatt, Dennis and Laura Klatt (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers". In: *The Journal of the Acoustical Society of America* 87 (2), pp. 820–857.

Kreiman, Jody, Norma Antonanzas-Barroso, and Bruce R. Gerratt (2016). "The UCLA voice synthesizer, version 2.0". In: *The Journal of the Acoustical Society of America* 140 (4), pp. 2961–2961.

Kreiman, Jody and Bruce Gerratt (2005). "Perception of aperiodicity in pathological voice". In: *The Journal of the Acoustical Society of America* 117 (4), pp. 2201–2211. eprint: `https://doi.org/10.1121/1.1858351`.

Kreiman, Jody, Bruce Gerratt, and Norma Antoñanzas-Barroso (2007). "Measures of the Glottal Source Spectrum". In: *Journal of Speech, Language, and Hearing Research* 50 (3), pp. 595–610.

Kreiman, Jody, Bruce Gerratt, Marc Garellek, Robin Samlan, and Zhaoyan Zhang (2014). "Toward a unified theory of voice production and perception". In: *Loquens* 1 (1), p. 009.

Krom, Guus de (1993). "A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals". In: *Journal of Speech, Language, and Hearing Research* 36 (2), pp. 254–266.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen (2017). "lmerTest Package: Tests in Linear Mixed Effects Models". In: *Journal of Statistical Software* 82 (13), pp. 1–26.

Lass, Norman J, Karen R Hughes, Melanie D Bowyer, Lucille T Waters, and Victoria T Bourne (1976). “Speaker sex identification from voiced, whispered, and filtered isolated vowels”. In: *Journal of the Acoustical Society of America.*

Laver, John (1968). “Voice quality and indexical information”. In: *British Jounral of Disorders of Communication* 3, pp. 43–54.

Laver, John (1973). “Labels for voices”. In: *Journal of the International Phonetic Association* 4, pp. 62–75.

Laver, John (1980). *The Phonetic Description of Voice Quality.* Cambridge: Cambridge University Press.

Lee, Kaitlyn (2016). “The Perception of Creaky Voice: Does Speaker Gender Affect our Judgments?” MA thesis. University of Kentucky.

Leung, Yeptain, Jennifer Oates, and Siew Pang Chan (2018). “Voice, Articulation, and Prosody Contribute to Listener Perceptions of Speaker Gender: A Systematic Review and Meta-Analysis”. In: *Journal of Speech, Language and Hearing Research* 61 (2), pp. 266–297.

Loveday, Leo (1981). “Pitch, Politeness and Sexual Role: An Exploratory Investigation into the Pitch Correlates of English and Japanese Politeness Formulae”. In: *Language and Speech* 24 (1), pp. 71–89.

Mason, Katherine (2018). “Gendered Embodiment”. In: *Handbook of the Sociology of Gender.* Ed. by Barbara J. Risman, Carissa M. Froyum, and William J. Scarborough. Cham: Springer International Publishing, pp. 95–107.

McElhinny, Bonnie (2014). “Theorizing Gender in Sociolinguistics and Linguistic Anthropology: Toward Effective Interventions in Gender Inequity”. In: *The Handbook of Language, Gender, and Sexuality.* Ed. by Miriam Meyerhoff Susan Ehrlich and Janet Holmes. 2nd edition. Chichester: Wiley Blackwell, pp. 48–66.

Moisik, Scott R., Míša Hejná, and John Esling (2019). “Abducted vocal fold states and the epilarynx: a new taxonomy for distinguishing breathiness and whisperiness”. In: *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019.* Ed. by Marija Tabain Sasha Calhoun Paola Escudero and Paul Warren. Canberra, Australia: Australasian Speech Science and Technology Association Inc., pp. 220–224.

Multi-dimensional voice program model 5105 (2004). Lincoln Park, USA.

Munson, Benjamin (2007). “The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation”. In: *Language and Speech.*

Oakley, Ann (2016[1972]). *Sex, gender, and society.* Abingdon: Routledge.

Ochs, Eleanor (1992). “Indexing gender”. In: *Rethinking Context: Language as an Interactive Phenomenon.* Ed. by A. Duranti and C. Goodwin. Cambridge: Cambridge University Press, pp. 335–358.

Owen, Kelly and Adrienne Hancock (2010). “The role of self- and listener perceptions of femininity in voice therapy”. In: *International Journal of Transgenderism* 12 (4), pp. 272–284.

Palmer, Derek, Angela Dietsch, and Jeff Searl (2012). “Endoscopic and stroboscopic presentation of the larynx in male-to-female transsexual persons”. In: *Journal of Voice* 26 (1), pp. 117–126.

Pearce, Jo (2019). “Identity, socialization and environment in transgender speakers: Sociophonetic variation in creak and /s/”. In: *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019.* Ed. by Marija Tabain Paul Warren Sasha Calhoun Paola Escudero. Canberra, Australia: Australasian Speech Science and Technology Association Inc., pp. 3290–3294.

Peirce, Charles S., Charles Hartshorne, Paul Weiss, and Arthur W. Burks (1931). *Collected papers.* English. Cambridge (Mass): Harvard U.P.

Pisanski, Katarzyna and Drew Rendall (2011). “The prioritization of voice fundamental frequency or formants in listeners’ assessments of speaker size, masculinity, and attractiveness”. In: *The Journal of the Acoustical Society of America* 129 (4), pp. 2201–2212.

Podesva, Robert (2007). “Phonation type as a stylistic variable: The use of falsetto in constructing a persona”. In: *Journal of Sociolinguistics* 11 (4), pp. 478–504. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9841.2007.00334.x`.

Podesva, Robert (2013). “Gender and the social meaning of non-modal phonation types”. In: *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society*, pp. 427–448.

Podesva, Robert and Patrick Callier (2015). “Voice quality and identity”. In: *Annual Review of Applied Linguistics* 35 (2015), pp. 173–194.

Podesva, Robert and Elaine Chun (2010). “Voice quality and indeterminacies of social meaning in constructed dialogue”. In: *Sociolinguistics Symposium 18.*

Podesva, Robert and Sakiko Kajino (2014). “Sociophonetics, Gender, and Sexuality”. In: *The Handbook of Language, Gender, and Sexuality.* John Wiley Sons, Ltd. Chap. 5, pp. 103–122. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118584248.ch5`.

Porter, Courtney Cain (2012). “Voice quality and gender identification: Acoustic and perceptual analysis”. PhD thesis. Dalhousie University.

Prolific (2014). *Prolific.* Version July 2019. Oxford.

R Core Team (2019). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria.

Rose, Phil (1991). “How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency?” In: *Speech Communication* 10 (3), pp. 229–247.

Royal College of Speech and Language Therapists (2018). *Trans and gender-diverse voice & communication therapy: Competency framework.* London: Royal College of Speech and Language Therapists.

Rubin, Gayle (1975). “The traffic in women: Notes on the “political economy” of sex”. In: *Toward an anthropology of women*. Ed. by Rayna R. Reiter. New York/London: Monthly Review Press.

Sawyer, Jean (2019). “Voice changes in transgender care”. In: *Transgender medicine: A multidisciplinary approach*. Ed. by Leonid Poretisky and Wylie C. Hembree. Cham: Springer, pp. 213–237.

Shue, Yen-Liang (2010). “The voice source in speech production: Data, analysis and models”. PhD thesis. UCLA.

Silverstein, Michael (1976). “Linguistic Categories and Cultural Description”. In: *Meaning in anthropology*. Alberquerque: University of New Mexico Press, pp. 11–55.

Silverstein, Michael (2003). “Indexical order and the dialectics of sociolinguistic life”. In: *Language Communication* 23 (3). Words and Beyond: Linguistic and Semiotic Studies of Sociocultural Order, pp. 193–229.

Simpson, Adrian (2009). “Phonetic differences between male and female speech female speech”. In: *Language and Linguistics Compass* 3 (2), pp. 621–640.

Skuk, Verena G and Stefan R Schweinberger (2014). “Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender”. In: *Journal of Speech Language and Hearing Research* 57 (1), p. 285.

Stevens, Kenneth (1977). “Physics of Laryngeal Behavior and Larynx Modes”. In: *Phonetica* 34, pp. 264–279.

Stoet, Gijsbert (2010). “PsyToolkit - A software package for programming psychological experiments using Linux”. In: *Behaviour Research Methods* 42 (4), pp. 1096–1104.

Stoet, Gijsbert (2017). “PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments”. In: *Teaching of Psychology* 44 (1), pp. 24–31.

Stuart-Smith, Jane (1999). “Voice quality in Glaswegian”. In: *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 2553–2556.

Syrdal, Ann K. (1996). “Acoustic variability in spontaneous conversational speech of American English talkers”. In: *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP ’96*, 438–441 vol.1.

Szakay, Anita (2006). “Rhythm and pitch as markers of ethnicity in New Zealand English”. In: *Proceedings of the 11th Australasian International Conference on Speech Science and Technology*. Ed. by Paul Warren and Catherine Watson. Canberra: Australasian Speech Science and Technology Association (ASSTA), pp. 421–426.

Szakay, Anita and Eivind Torgersen (2015). “An Acoustic Analysis of Voice Quality in London English: the Effect of Gender, Ethnicity and F0”. In: *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*.

Szakay, Anita and Eivind Torgersen (2019). “A re-analysis of f0 in ethnic varieties of London English using REAPER”. In: *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Ed. by Marija Tabain Sasha Calhoun Paola Escudero and Paul Warren. Canberra, Australia: Australasian Speech Science and Technology Association Inc., pp. 1675–1678.

Talkin, David (2015). *REAPER: Robust Epoch And Pitch EstimatoR.* URL: `https://github.com/google/REAPER`.

Titze, Ingo R (Apr. 1989). "Physiologic and acoustic differences between male and female voices". In: *The Journal of the Acoustical Society of America* 85 (4), pp. 1699–1707.

Traunmüller, Hartmut and Anders Eriksson (1995). "The frequency range of the voice fundamental in the speech of male and female adults". Stockholm.

Van Borsel, John, Joke Janssens, and Marc De Bodt (2009). "Breathiness as a feminine voice characteristic: A perceptual approach". In: *Journal of Voice* 23 (3), pp. 291–294.

Van Dyke, Henry (1918). *Muscles of the larynx, seen from above.* in Gray (1918).

Weenink, David (2009). *The KlattGrid speech synthesizer.* Amsterdam.

West, Candace and Don H. Zimmerman (1987). "Doing gender". In: *Gender and Society* 1 (2), pp. 125–151.

Whiteside, S. P. (1998). "Identification of a Speaker's Sex: A Study of Vowels". In: *Perceptual and Motor Skills* 86 (2), pp. 579–584.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.

Wolfe, Virginia I., David L. Ratusnik, Furman H. Smith, and Gretajo Northrop (1990). "Intonation and Fundamental Frequency in Male-to-Female Transsexuals". In: *Journal of Speech and Hearing Disorders* 55 (1), pp. 43–50.

Wolk, Lesley, Nassima B. Abdelli-Beruh, and Dianne Slavin (2012). "Habitual use of vocal fry in young adult female speakers". In: *Journal of Voice* 26 (3), e111–e116.

Yen-Liang Shue Patricia Keating, Chad Vicenik Kristine Yu (2011). "VoiceSauce: A program for voice analysis". In: *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)*, pp. 1846–1849.

Yuasa, Ikuko (2008). *Culture and gender of voice pitch: A sociophonetic comparison of the Japanese and Americans.* London: Equinox.

Yuasa, Ikuko (2010). "Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women?" In: *American Speech* 85 (3), pp. 315–337.

Yumoto, E., W. J. Gould, and T. Baer (1982). "Harmonics-to-noise ratio as an index of the degree of hoarseness". English. In: *The Journal of the Acoustical Society of America* 71 (6), pp. 1544–1550.

Zimman, Lal (2013). "Hegemonic masculinity and the variability of gay-sounding speech: The perceived sexuality of transgender men". In: *Journal of Language and Sexuality* 2 (1), pp. 1–39.

Zimman, Lal (2018). "Transgender voices: Insights on identity, embodiment, and the gender of the voice". In: *Language and Linguistics Compass* 12 (8). e12284 LNCO-

0742, e12284. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/lnc3.12284.

Appendix A

Praat script used for KlattGrid copy synthesis and voice quality manipulations

```
# This Praat script will open all the .wav files in a directory, then go
through each file and for each one create three new Praat objects and
alter the median F0 level of each, and use KlattGrid copy synthesis to
produce sound objects modal, breathy and creaky voice. The user then has
the option to save these sound objects in a directory.


# This part of the script makes a form pop up, so that the user can enter
what settings to use and what they want Praat to change


form Change pitch and voice quality

# KlattGrid copy synthesis needs to know the ceiling of the formant
search range - check the Praat manual for more information, but as a
baseline make maxformant 5000Hz for female speakers and 5500Hz for male
speakers

real maxformant 5000

comment Which new f0s do you want?
real newpitch1 120
real newpitch2 165
real newpitch3 210
```

```
comment Modal voice parametres
boolean make_modal 1
real flutter_modal 0.1
real open_phase_modal 0.7
real spectral_tilt_modal 10

comment Creaky voice parametres
boolean make_creaky 1
real flutter_creak 0.25
real open_phase_creak 0.4
real spectral_tilt_creak 5
real max_double_pulsing 0.25

comment Breathy voice parametres
boolean make_breathy 1
real flutter_breathy 0.25
real b1_increase_breathy 1.10 (=10%)
real open_phase_breathy 0.95
real spectral_tilt_breathy 30

comment Choose directories to open and save files in
word opendirectory N:\originalstimulitest\
boolean save_files 1
word savedirectory N:\editedstimulitest\

endform

strings = Create Strings as file list: "list", opendirectory$ + "*.wav"
numberOfFiles = Get number of strings
for ifile to numberOfFiles
    selectObject: strings
    fileName$ = Get string: ifile
   Read from file: opendirectory$ + fileName$

soundID = selected ("Sound")
selectObject: soundID
soundName$ = selected$ ("Sound")


#getting time measurements
```

```
tmin = Get start time
tmax = Get end time

#changing pitch



selectObject: soundID
Change gender: 75, 600, 1, newpitch1, 1, 1
Rename: soundName$ + "_" + string$ (newpitch1)
soundID1 = selected ("Sound")

selectObject: soundID
Change gender: 75, 600, 1, newpitch2, 1, 1
Rename: soundName$ + "_" +  string$ (newpitch2)
soundID2 = selected ("Sound")

selectObject: soundID
Change gender: 75, 600, 1, newpitch3, 1, 1
Rename: soundName$ + "_" +  string$ (newpitch3)
soundID3 = selected ("Sound")

#loop that goes through sounds and creates new vq sounds for each of them

for isound from soundID1 to soundID3

selectObject: isound
name$ = selected$ ("Sound")

To Pitch: 0, 75, 600

minpitch = Get minimum: 0, 0, "Hertz", "Parabolic"
maxpitch = Get maximum: 0, 0, "Hertz", "Parabolic"
q1pitch = Get quantile: 0, 0, 0.125, "Hertz"
q2pitch = Get quantile: 0, 0, 0.25, "Hertz"
q3pitch = Get quantile: 0, 0, 0.375, "Hertz"
medianpitch = Get quantile: 0, 0, 0.5, "Hertz"

Remove

#modal sample
#transforming the sound to a KlattGrid, setting parametres,
#convert back to sound and rename

if make_modal = 1

selectObject: isound
To KlattGrid (simple): 0.005, 5, maxformant, 0.025, 50, 60, 600, 100, "yes"
Add flutter point: 0, flutter_modal
Add open phase point: 0, open_phase_modal
Add spectral tilt point: 0, spectral_tilt_modal
To Sound
newname$ = name$ + "modal"
Rename: name$ + "modal"

if save_files = 1
Save as WAV file: savedirectory$ + newname$ + ".wav"
endif

selectObject: "KlattGrid " + name$
Remove

endif

#creaky sample

if make_creaky = 1

selectObject: isound


To KlattGrid (simple): 0.005, 5, maxformant, 0.025, 50, 60, 600, 100, "yes"

Add open phase point: 0, open_phase_creak
Add spectral tilt point: 0, spectral_tilt_creak
Add flutter point: 0, flutter_creak

for i to (tmax-tmin)/0.01
time = tmin + i * 0.01
pitch = Get pitch at time: time
if pitch < q1pitch
Remove pitch points: time - 0.005, time + 0.005
Add pitch point: time, pitch - 12
Add double pulsing point: time, max_double_pulsing
elsif pitch < q2pitch
Remove pitch points: time - 0.005, time + 0.005
Add pitch point: time, pitch - 9
Add double pulsing point: time, max_double_pulsing
elsif pitch < q3pitch
Remove pitch points: time - 0.005, time + 0.005
Add pitch point: time, pitch - 6
Add double pulsing point: time, max_double_pulsing
elsif pitch < medianpitch
Remove pitch points: time - 0.005, time + 0.005
Add pitch point: time, pitch - 3
Add double pulsing point: time, max_double_pulsing/2
else
Add double pulsing point: time, 0.0
endif
endfor

To Sound
newname$ = name$ + "creaky"
Rename: name$ + "creaky"

if save_files = 1
Save as WAV file: savedirectory$ + newname$ + ".wav"
endif

selectObject: "KlattGrid " + name$
Remove
endif

# breathy


if make_breathy = 1

selectObject: isound

To KlattGrid (simple): 0.005, 5, maxformant, 0.025, 50, 60, 600, 100, "yes"

for i to (tmax-tmin)/0.01
time = tmin + i * 0.01
```

```
for i to (tmax-tmin)/0.01
time = tmin + i * 0.01
b1 = Get oral formant bandwidth at time: 1, tmin + i * 0.01
Remove oral formant bandwidth points: 1, time, time
Add oral formant bandwidth point: 1, time, b1*b1_increase_breathy
endfor

for i to (tmax-tmin)/0.01
time = tmin + i * 0.01
intensity = Get voicing amplitude at time: time
Add aspiration amplitude point: time, intensity/2.5
endfor



endfor




Add open phase point: 0, open_phase_breathy

Add spectral tilt point: 0, spectral_tilt_breathy

Add flutter point: 0, flutter_breathy

To Sound
newname$ = name$ + "breathy"
Rename: name$ + "breathy"
if save_files = 1
Save as WAV file: savedirectory$ + newname$ + ".wav"
endif
selectObject: "KlattGrid " + name$
Remove

endif

endfor

selectObject: soundID
Remove
selectObject: soundID1
```

```
Remove
selectObject: soundID2
Remove
selectObject: soundID3
Remove

endfor
```

Appendix B

List of phrases and sentences used for stimuli

Breathy

I'll wear wool
I rely on you
I owe you money
I'm lying
We mean well
I know a lawyer
You're in Norway
I know you're away
I'll allow you a year
I'm earning money
Our lawyer won
You're alone now
You're wearing a wire
I'm ironing linen
I ran a long way
You're normally early
I'm really lonely
I'm marrying you
An unruly manner
A real whale
A lowly warrior
A real worry
A lone warrior
A rainy morning
A long year
A long lorry
A rural area
A rainy year
A new moon
A new year
A long memoir
A warm year
A rare eel
A wee lamb
A wirey willow
A wooly ewe

Creaky

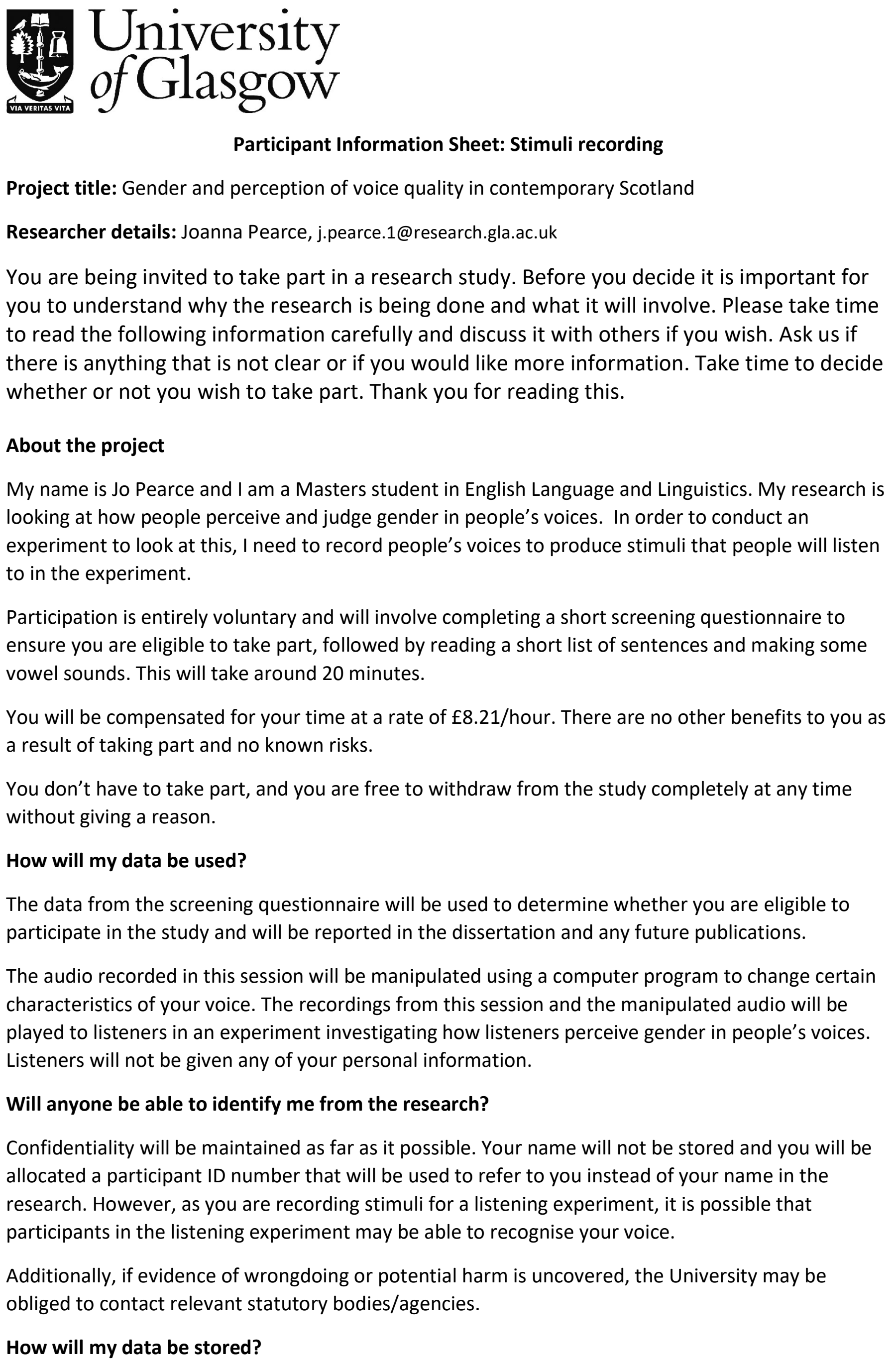
I'm in Romania
I'm rarely wrong
You're a memory
You're a miner
I wore a lily
We're in a union
I'm on my way
I'm near a railway line
You're a liar
I'm running a mile
I'll lure you away
We were running
You're really wrong
You're yawning
We were lying
You know I worry
I know you'll learn
I normally win
A low wall
A yellow lily
A narrow meaning
A weary ruler
A long year
A yellow lorry
A loyal ruler
A long wall
A wee lie
A Roman warrior
An annual renewal
A mile away
A lying lawyer
A knowing owl
A young lion
A reigning royal
A rare owl
A young lawyer

Modal

You're wearing a ring
I know you
We're in Iran
I'm a royal
A wirey willow
I know you well
A weary lawyer
I rang you
I'm learning
I'm near you
You're loyal
We'll wear yellow
You're yelling
You're worrying me
You're really ill
You're annoying
You won money
I know where you are
A wee while
A yellow eye
A yearly reunion
A warm morning
A willing ruler
A wooly lamb
A long way
A yellow welly
A yellow lion
A lying lawyer
A new law
A new name
A narrow alley
A long railway
A normal morning
A wee lion
A weary lawyer
A young miner

Appendix C

Participant information sheet



University of Glasgow

Participant Information Sheet: Stimuli recording

Project title: Gender and perception of voice quality in contemporary Scotland

Researcher details: Joanna Pearce, j.pearce.1@research.gla.ac.uk

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

About the project

My name is Jo Pearce and I am a Masters student in English Language and Linguistics. My research is looking at how people perceive and judge gender in people's voices. In order to conduct an experiment to look at this, I need to record people's voices to produce stimuli that people will listen to in the experiment.

Participation is entirely voluntary and will involve completing a short screening questionnaire to ensure you are eligible to take part, followed by reading a short list of sentences and making some vowel sounds. This will take around 20 minutes.

You will be compensated for your time at a rate of £8.21/hour. There are no other benefits to you as a result of taking part and no known risks.

You don't have to take part, and you are free to withdraw from the study completely at any time without giving a reason.

How will my data be used?

The data from the screening questionnaire will be used to determine whether you are eligible to participate in the study and will be reported in the dissertation and any future publications.

The audio recorded in this session will be manipulated using a computer program to change certain characteristics of your voice. The recordings from this session and the manipulated audio will be played to listeners in an experiment investigating how listeners perceive gender in people's voices. Listeners will not be given any of your personal information.

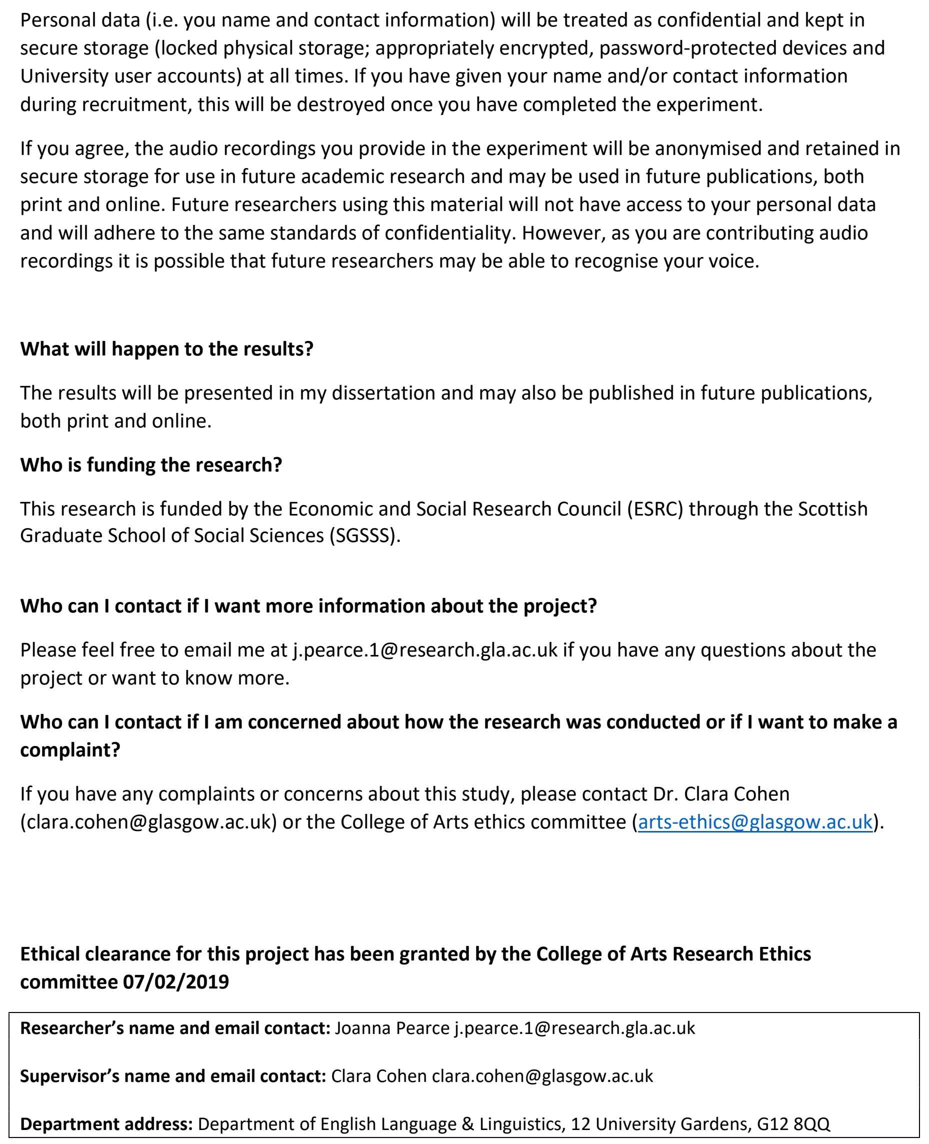
Will anyone be able to identify me from the research?

Confidentiality will be maintained as far as it possible. Your name will not be stored and you will be allocated a participant ID number that will be used to refer to you instead of your name in the research. However, as you are recording stimuli for a listening experiment, it is possible that participants in the listening experiment may be able to recognise your voice.

Additionally, if evidence of wrongdoing or potential harm is uncovered, the University may be obliged to contact relevant statutory bodies/agencies.

How will my data be stored?

Figure C.1: Front side of participant information sheet used for stimuli recording sessions



Personal data (i.e. you name and contact information) will be treated as confidential and kept in secure storage (locked physical storage; appropriately encrypted, password-protected devices and University user accounts) at all times. If you have given your name and/or contact information during recruitment, this will be destroyed once you have completed the experiment.

If you agree, the audio recordings you provide in the experiment will be anonymised and retained in secure storage for use in future academic research and may be used in future publications, both print and online. Future researchers using this material will not have access to your personal data and will adhere to the same standards of confidentiality. However, as you are contributing audio recordings it is possible that future researchers may be able to recognise your voice.

What will happen to the results?

The results will be presented in my dissertation and may also be published in future publications, both print and online.

Who is funding the research?

This research is funded by the Economic and Social Research Council (ESRC) through the Scottish Graduate School of Social Sciences (SGSSS).

Who can I contact if I want more information about the project?

Please feel free to email me at j.pearce.1@research.gla.ac.uk if you have any questions about the project or want to know more.

Who can I contact if I am concerned about how the research was conducted or if I want to make a complaint?

If you have any complaints or concerns about this study, please contact Dr. Clara Cohen (clara.cohen@glasgow.ac.uk) or the College of Arts ethics committee (arts-ethics@glasgow.ac.uk).

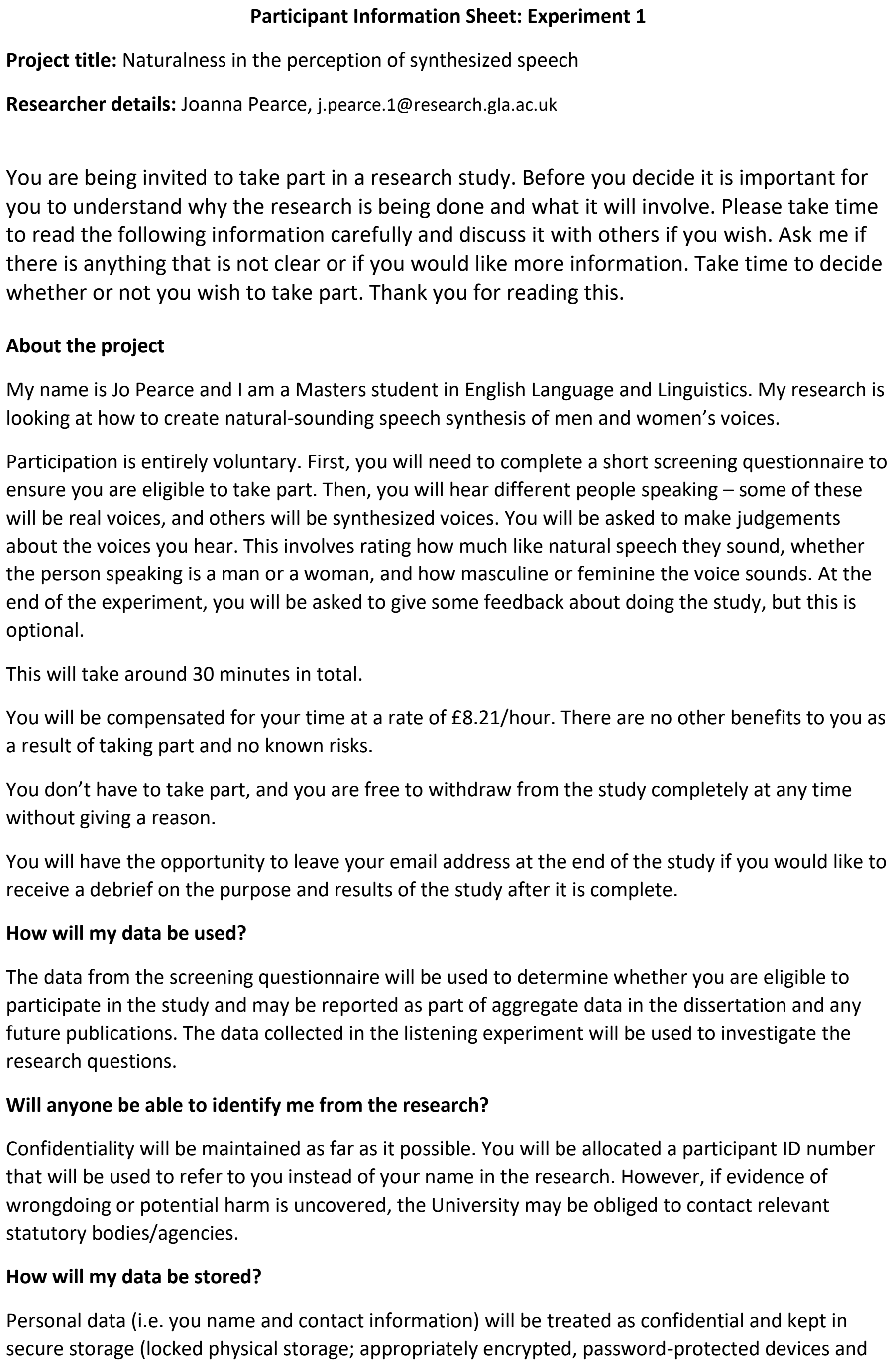
Ethical clearance for this project has been granted by the College of Arts Research Ethics committee 07/02/2019

Researcher's name and email contact: Joanna Pearce j.pearce.1@research.gla.ac.uk

Supervisor's name and email contact: Clara Cohen clara.cohen@glasgow.ac.uk

Department address: Department of English Language & Linguistics, 12 University Gardens, G12 8QQ

Figure C.2: Back side of participant information sheet used for stimuli recording sessions



Participant Information Sheet: Experiment 1

Project title: Naturalness in the perception of synthesized speech

Researcher details: Joanna Pearce, j.pearce.1@research.gla.ac.uk

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

About the project

My name is Jo Pearce and I am a Masters student in English Language and Linguistics. My research is looking at how to create natural-sounding speech synthesis of men and women's voices.

Participation is entirely voluntary. First, you will need to complete a short screening questionnaire to ensure you are eligible to take part. Then, you will hear different people speaking – some of these will be real voices, and others will be synthesized voices. You will be asked to make judgements about the voices you hear. This involves rating how much like natural speech they sound, whether the person speaking is a man or a woman, and how masculine or feminine the voice sounds. At the end of the experiment, you will be asked to give some feedback about doing the study, but this is optional.

This will take around 30 minutes in total.

You will be compensated for your time at a rate of £8.21/hour. There are no other benefits to you as a result of taking part and no known risks.

You don't have to take part, and you are free to withdraw from the study completely at any time without giving a reason.

You will have the opportunity to leave your email address at the end of the study if you would like to receive a debrief on the purpose and results of the study after it is complete.

How will my data be used?

The data from the screening questionnaire will be used to determine whether you are eligible to participate in the study and may be reported as part of aggregate data in the dissertation and any future publications. The data collected in the listening experiment will be used to investigate the research questions.

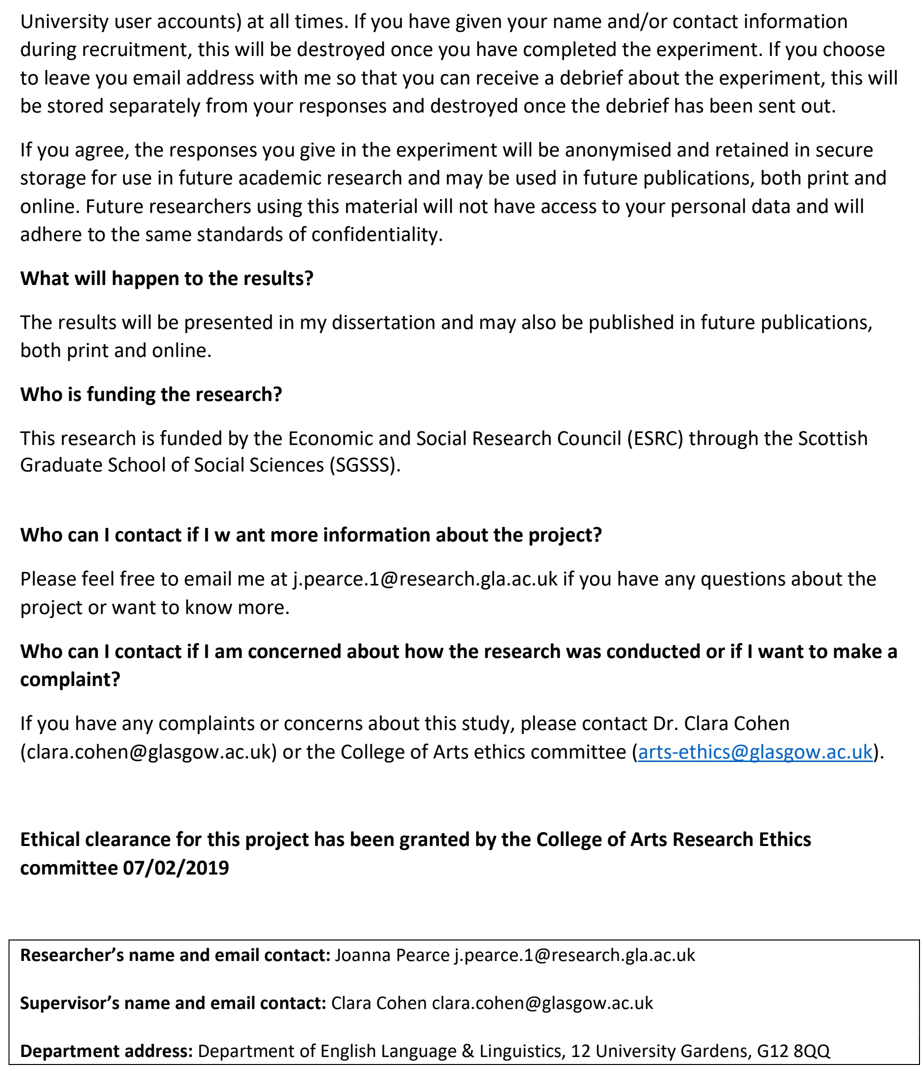
Will anyone be able to identify me from the research?

Confidentiality will be maintained as far as it possible. You will be allocated a participant ID number that will be used to refer to you instead of your name in the research. However, if evidence of wrongdoing or potential harm is uncovered, the University may be obliged to contact relevant statutory bodies/agencies.

How will my data be stored?

Personal data (i.e. you name and contact information) will be treated as confidential and kept in secure storage (locked physical storage; appropriately encrypted, password-protected devices and

Figure C.3: Front side of participant information sheet used for Experiment 1



University user accounts) at all times. If you have given your name and/or contact information during recruitment, this will be destroyed once you have completed the experiment. If you choose to leave you email address with me so that you can receive a debrief about the experiment, this will be stored separately from your responses and destroyed once the debrief has been sent out.

If you agree, the responses you give in the experiment will be anonymised and retained in secure storage for use in future academic research and may be used in future publications, both print and online. Future researchers using this material will not have access to your personal data and will adhere to the same standards of confidentiality.

What will happen to the results?

The results will be presented in my dissertation and may also be published in future publications, both print and online.

Who is funding the research?

This research is funded by the Economic and Social Research Council (ESRC) through the Scottish Graduate School of Social Sciences (SGSSS).

Who can I contact if I w ant more information about the project?

Please feel free to email me at j.pearce.1@research.gla.ac.uk if you have any questions about the project or want to know more.

Who can I contact if I am concerned about how the research was conducted or if I want to make a complaint?

If you have any complaints or concerns about this study, please contact Dr. Clara Cohen (clara.cohen@glasgow.ac.uk) or the College of Arts ethics committee (arts-ethics@glasgow.ac.uk).

Ethical clearance for this project has been granted by the College of Arts Research Ethics committee 07/02/2019

Researcher's name and email contact: Joanna Pearce j.pearce.1@research.gla.ac.uk

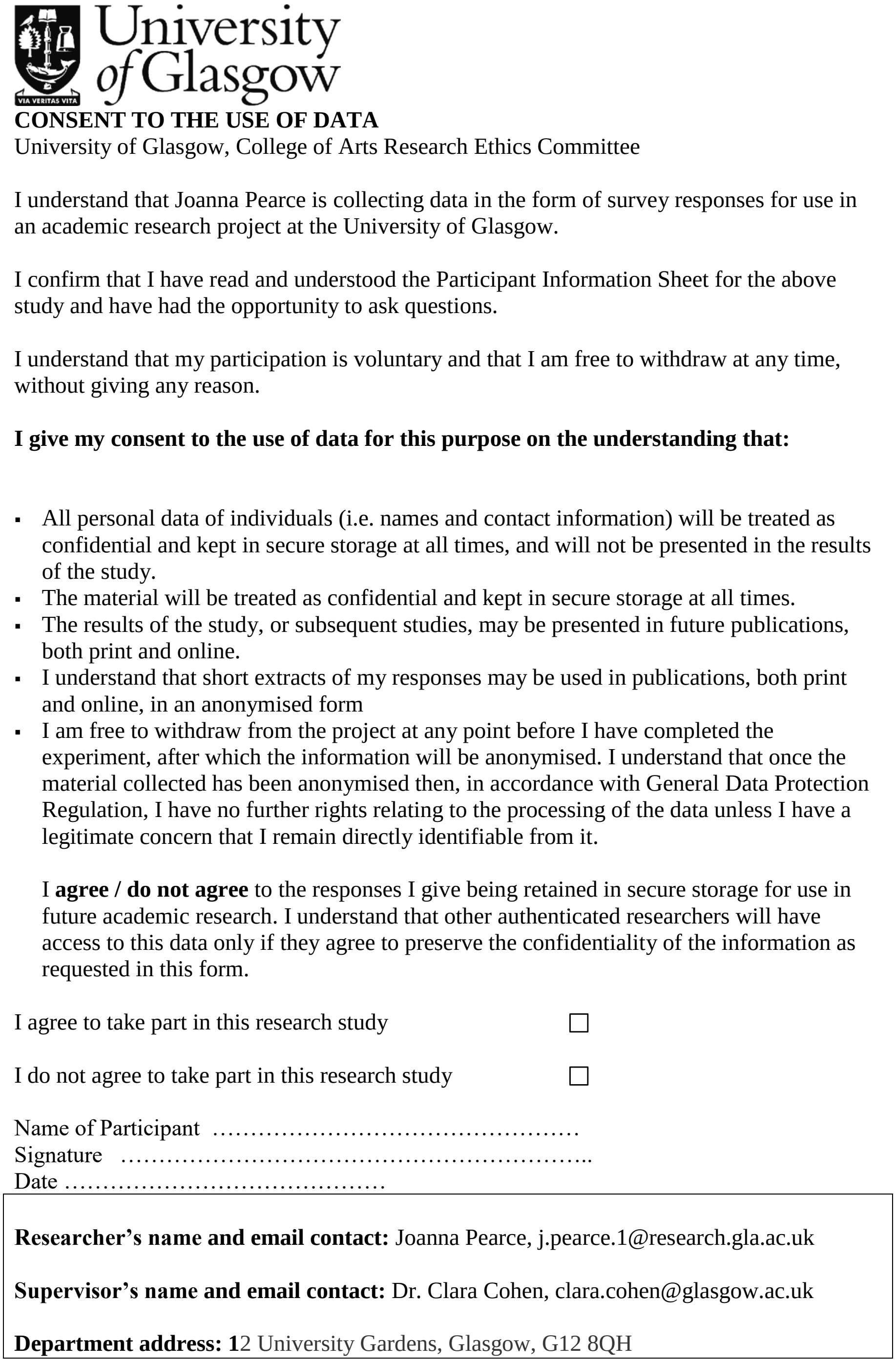
Supervisor's name and email contact: Clara Cohen clara.cohen@glasgow.ac.uk

Department address: Department of English Language & Linguistics, 12 University Gardens, G12 8QQ

Figure C.4: Back side of participant information sheet used for Experiment 1

Appendix D

Consent form



CONSENT TO THE USE OF DATA
University of Glasgow, College of Arts Research Ethics Committee

I understand that Joanna Pearce is collecting data in the form of survey responses for use in an academic research project at the University of Glasgow.

I confirm that I have read and understood the Participant Information Sheet for the above study and have had the opportunity to ask questions.

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

I give my consent to the use of data for this purpose on the understanding that:

- All personal data of individuals (i.e. names and contact information) will be treated as confidential and kept in secure storage at all times, and will not be presented in the results of the study.
- The material will be treated as confidential and kept in secure storage at all times.
- The results of the study, or subsequent studies, may be presented in future publications, both print and online.
- I understand that short extracts of my responses may be used in publications, both print and online, in an anonymised form
- I am free to withdraw from the project at any point before I have completed the experiment, after which the information will be anonymised. I understand that once the material collected has been anonymised then, in accordance with General Data Protection Regulation, I have no further rights relating to the processing of the data unless I have a legitimate concern that I remain directly identifiable from it.

  I **agree / do not agree** to the responses I give being retained in secure storage for use in future academic research. I understand that other authenticated researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.

I agree to take part in this research study □

I do not agree to take part in this research study □

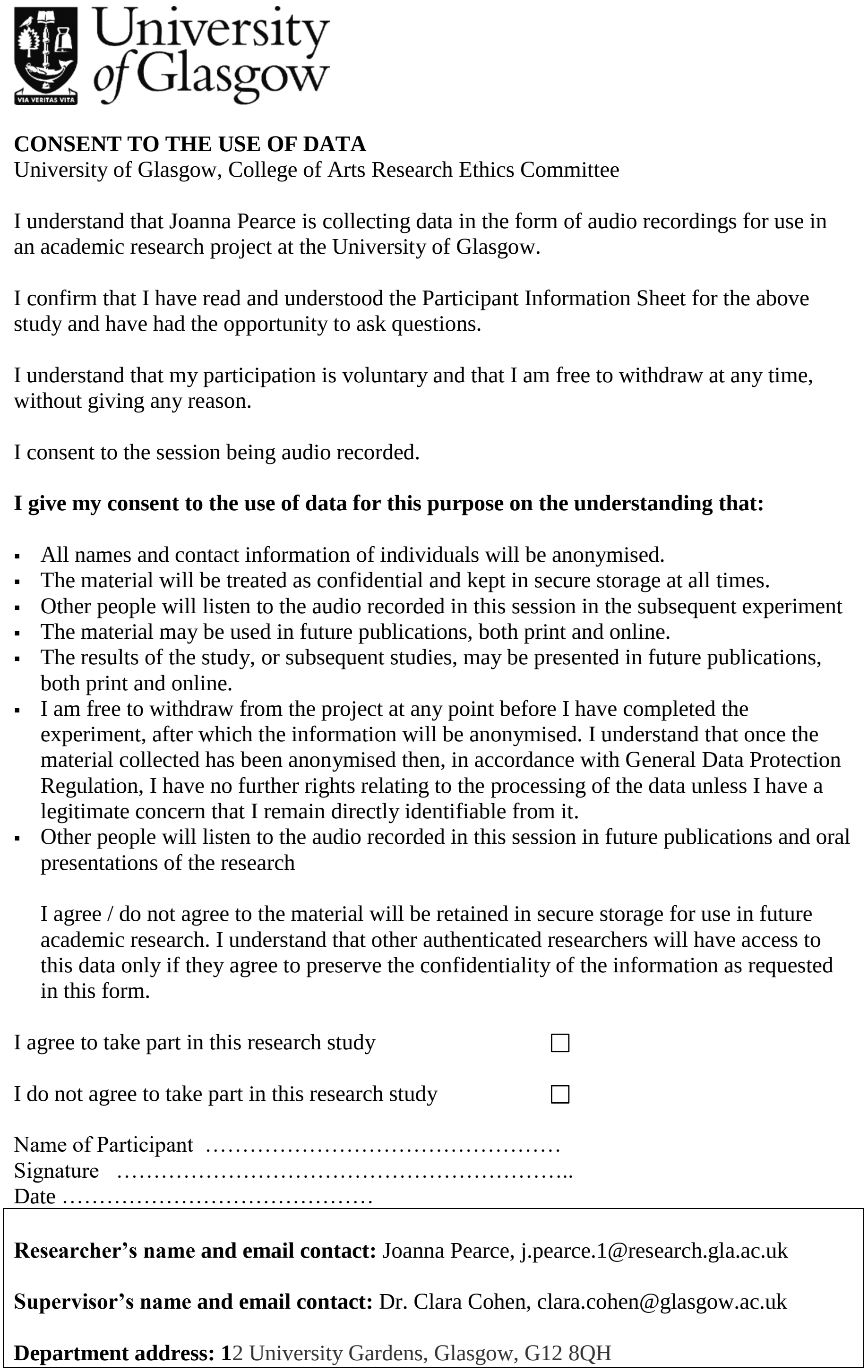
Name of Participant …………………………………………
Signature …………………………………………………..
Date ……………………………………

Researcher's name and email contact: Joanna Pearce, j.pearce.1@research.gla.ac.uk

Supervisor's name and email contact: Dr. Clara Cohen, clara.cohen@glasgow.ac.uk

Department address: 12 University Gardens, Glasgow, G12 8QH

Figure D.1: Consent form used for Experiment 1



University of Glasgow

CONSENT TO THE USE OF DATA
University of Glasgow, College of Arts Research Ethics Committee

I understand that Joanna Pearce is collecting data in the form of audio recordings for use in an academic research project at the University of Glasgow.

I confirm that I have read and understood the Participant Information Sheet for the above study and have had the opportunity to ask questions.

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

I consent to the session being audio recorded.

I give my consent to the use of data for this purpose on the understanding that:

- All names and contact information of individuals will be anonymised.
- The material will be treated as confidential and kept in secure storage at all times.
- Other people will listen to the audio recorded in this session in the subsequent experiment
- The material may be used in future publications, both print and online.
- The results of the study, or subsequent studies, may be presented in future publications, both print and online.
- I am free to withdraw from the project at any point before I have completed the experiment, after which the information will be anonymised. I understand that once the material collected has been anonymised then, in accordance with General Data Protection Regulation, I have no further rights relating to the processing of the data unless I have a legitimate concern that I remain directly identifiable from it.
- Other people will listen to the audio recorded in this session in future publications and oral presentations of the research

I agree / do not agree to the material will be retained in secure storage for use in future academic research. I understand that other authenticated researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.

I agree to take part in this research study □

I do not agree to take part in this research study □

Name of Participant …………………………………………
Signature ……………………………………………………..
Date ……………………………………

Researcher's name and email contact: Joanna Pearce, j.pearce.1@research.gla.ac.uk

Supervisor's name and email contact: Dr. Clara Cohen, clara.cohen@glasgow.ac.uk

Department address: 12 University Gardens, Glasgow, G12 8QH

Figure D.2: Consent form used for stimuli recording sessions