Social meaning in Scottish voices

A sociophonetic investigation of voice quality combining auditory-perceptual, acoustic, and qualitative approaches

Joe Pearce

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF CRITICAL STUDIES

COLLEGE OF ARTS

MAY 2024

Abstract

Voice quality – the characteristic features of a person's voice – varies according to macro-social categories like social class, age and gender (Esling 1978b, Stuart-Smith 1999b, Beck & Schaeffler 2015), and within speakers as they present different personae to others (Podesva 2007). Here, I focus on laryngeal aspects of voice quality, also called phonation, and consider how combining auditory-perceptual, acoustic, and qualitative methods can enrich the study of voice quality and its social meaning. I take the context of Scottish voice quality, and ask how this varies by age, gender and region, and then use Interpretative Phenomenological Analysis (IPA) to consider how these meanings are constructed and understood by an individual speaker.

I first considered the relationships between acoustic measures and descriptive voice quality labels, with a view to evaluating the feasibility of conducting an acoustic analysis at scale in spontaneous speech. 90 seconds of speech from 24 speakers of Glasgow, Lothian (in and around Edinburgh) and Insular Scots from the Scots Syntax Atlas (SCOSYA) (Smith et al. 2019), stratified by age (18-25 and 65+) and gender (male and female), were investigated in a linked auditory-perceptual and acoustic study. Phonation Profile Analysis (PPA), a novel method that takes the descriptive phonation labels from Vocal Profile Analysis (Laver et al. 1991[1981]) and applies them to short stretches of sonorants, revealed a prevalence of whispery and tense-whispery voice. PPA results for 2170 stretches were compared to automatic f0-based categorization of creak (Dallaston & Docherty 2019) and acoustic analysis of H1*-H2*, H2*-H4* and H4*-2kHz* and Cepstral Peak Prominence (CPP) taken using VoiceSauce (Shue et al. 2011) informed by the psychoacoustic model (Kreiman et al. 2021). I concluded that it would be possible to use these acoustic methods at scale and maintain interpretability of results: f0-based categorisation of creak and PPA coding of creak showed high agreement, and whispery, breathy, tense, modal, and tense-whispery voice showed different acoustic profiles each on at least one acoustic measure, aiding in the interpretation of later corpus results.

I then used these methods with 180 seconds of speech from 95 male and female, older and younger speakers from SCOSYA (Glasgow n=48, Lothian n=28, Insular n=19). In contrast to previous findings by Stuart-Smith (1999b) and Beck & Schaeffler (2015), I found no gender difference in the use of creak, instead finding that Insular Scots speakers were particularly creaky. Analysis of spectral tilt and CPP revealed that female speakers used more tense phonation than male speakers, while younger speakers used a tenser and more near-modal quality than older speakers.

I then combined this acoustic approach with Interpretative Phenomenological Analysis (IPA), a qualitative approach interested in how people understand major life experiences, in a case study of how Carrie, a Scottish transgender woman, uses her voice

and understands her experience with it. Experiential themes highlight the importance of situational control in how she uses her voice: Carrie understands others to perceive a disconnect between her appearance and her voice, and while she expresses feeling self-conscious about her voice and feeling a need to change it for others in everyday situations, she takes joy in using her voice professionally and exploiting the perceived disconnect her voice creates in situations where she is in control. She uses harsh voice to voice how she believes others see this disconnect, drawing on iconised links between harsh voice and wider societal attitudes towards trans women as monstruous.

Through combining auditory-perceptual analysis, automatic categorisation of creak, multiple measures of voice quality, and in-depth qualitative data, this research gives a fuller picture of view of social meaning in voice quality, by not only tracing how multiple aspects of voice quality vary according to macro-level social factors for Scottish speakers, and what one Scottish speaker's own use of voice quality means to her, and where her understanding of her voice fits into the wider landscape of social meaning in Scottish voices.

Acknowledgements

I would first of all like to thank Jane Stuart-Smith, Clara Cohen and Felix Schaeffler for their support and guidance through this process. This thesis would not exist without Jane's research on voice quality in Glasgow, and the passion for sociophonetics that she passed in the sociolinguistics and phonetics lectures she gave during my undergraduate degree. It also probably would never have been finished without Clara, who pulled me out of many rabbit holes.

I would like to thank Florence Oulds, Kirstie English and Arts Ethics Board for their input on various aspects of this research. I'd also like to thank everyone from Arts IT Support, the College of Arts Graduate School, and the University's Data Management Team who helped me with many practical issues.

I would like to thank all my fellow GULPers, who were an endless source of support through every stage of the process, especially to Julia Moreno, Ebtehal Al-Asiri and Fabienne Westerberg, who helped welcome me into the lab, as well as to Divyanshi Shaktawat, Margie Ferguson, and Edward Marshall, who were there all along the way. Finally, thank you to Nate Haj Bakir, Lucy Jackson, Ryan Shaw-Hawkins and all the other GULPers still toiling away on their theses - and of course, good luck!

Many thanks to all the friends who kept me sane throughout the PhD, particularly Jenny, Fin, Emily, Gracie and Jo. Another special thank you to Ed, who provided me with soup, much-needed hugs, and commiseration. I'm eternally grateful to Jemma, who got me through the hardest moments over the past four years and who fills my life with so much joy.

Finally, I'm endlessly grateful to Carrie and Eilidh, who lent their time and voices to this research, as well as to my pilot participants.

This work was supported by the Economic and Social Research Council (Grant number: 2178789).

List of Tables

List of Figures

Contents

Part I

Introduction

This thesis investigates the social meaning of voice quality in Scottish accents through a combination of auditory-perceptual, acoustic and qualitative approaches. Following Eckert & Labov (2017) and Hall-Lew, Moore & Podesva (2021: 3-4), I understand the 'social meaning' of voice quality to be the context-dependent set of inferences that a listener might make on the basis of how voice quality is used in an interaction, which might include inferences about the type of person who is speaking, what social groupings they belong to, the qualities they might possess, and the stances they are taking in the interaction. I focus on laryngeal voice quality, defined here as the role of phonation in creating the characteristic sound of a speaker's voice, drawing on Abercrombie (1967), Laver (1980) and Esling et al. (2019).

Previous research on Scottish voice quality has primarily taken an auditory-perceptual approach using Vocal Profile Analysis (VPA) (Laver et al. 1991[1981]) investigate how long-term vocal tract settings pattern according to age, gender and social class, in Edinburgh (Esling 1978b, Beck 1988), Glasgow (Stuart-Smith 1999a), Inverness, Aberdeen and Dumfries (Beck & Schaeffler 2015). This approach uses descriptive labels for voice quality that relate to Laver's (1980) systematic description of voice quality, allowing this research to reveal that Scottish accents are characterised by whispery voice (Beck 1988, Stuart-Smith 1999a, Beck & Schaeffler 2015), and that non-modal qualities such as harsh voice and creak vary according to social factors such as age, gender and social class (Esling 1978b, Beck 1988, Stuart-Smith 1999a, Beck & Schaeffler 2015). However, acoustic analysis of laryngeal voice quality that draws from psychoacoustically-validated measures (Kreiman et al. 2021, Gittelson, Leemann & Tomaschek 2021) may be useful for considering both social and linguistic factors in larger-scale corpus investigations of voice quality, while in-depth qualitative case studies (Podesva 2007) have shown promise for considering how social meaning might become attached to variation in voice quality. In this thesis, I aim to bring together these different approaches to consider social meaning in Scottish laryngeal voice quality.

In this thesis, I begin with a macro-level investigation of sociophonetic variation in voice quality, considering the following research question in Part II:

- How does Scottish laryngeal voice quality vary by age, gender and area?

As there is little sociophonetically-oriented acoustic research on Scottish voice quality, I aim first to consider the feasibility of conducting larger-scale, corpus-based acoustic research on voice quality. Psychoacoustic work on how listeners differentiate between samples of voices finds that perception of voices is multidimensional (Kreiman et al. 2021), and stresses that instead of relying on a single measure, combining measures allows both noise and glottal tension to be considered and allows clearer interpretation of results Garellek (2019). Previous work (Wileman 2018, Schaeffler, Eichner &

Beck 2019) begins to connect multiple acoustic measures to auditory-perceptual labels, but does not consider the acoustic manifestation of combination types, such as tense whispery voice, and auditory degree of a particular quality. Using data from 24 speakers from the SCOSYA corpus (Smith et al. 2019), I conducted a combined auditory-perceptual and acoustic study of voice quality in 90 seconds of speech from speakers from three areas, Glasgow, Lothian (Edinburgh and surrounding areas) and Shetland, to consider the following research questions:

- What is the auditory-perceptual profile that characterises laryngeal voice quality in Scottish accents?
- How can we use acoustic measures to investigate variation in voice quality in corpus data, while maintaining the ability to interpret the acoustic nature of voice quality in terms of auditory quality?

This part of the investigation establishes a methodology for considering multiple dimensions of voice quality in corpus research, laying the groundwork for further investigation. Furthermore, it demonstrates the importance of taking community-level norms and variability into account when interpreting acoustic results.

I then scaled up the acoustic analysis to a wider sample from SCOSYA to examine how voice quality varied according to social and linguistic factors, considering 180 seconds of speech from 95 speakers from Glasgow, Lothian (Edinburgh and surrounding areas) and both Orkney and Shetland (Insular Scots), stratified by gender (49 female, 46 male), age (49 aged 18-25, 46 aged 65+) and area (19 Insular, 28 Lothian, 48 Glasgow). I considered the following research question:

- How does Scottish voice quality vary acoustically according to social and linguistic factors?

This part of the research aimed to establish how voice quality in Scottish accents varied according to macro-level social factors.

I then considered how social meaning emerges in a case study of a single speaker's voice and experiences. Trans speakers are often well-placed to consider how social meaning is attributed to their voices: Previous research has found that trans speakers' voices are re-contextualised by others as they transition (Zimman 2017a), and that trans speakers often document changes in their voices and reflect on how changes in their voice affect how they are treated by others (Crowley 2021). I therefore chose to focus on a case study of a Scottish transgender speaker, Carrie, who uses her voice extensively in her professional life, as well as to navigate everyday situations such as telephone conversations and public transport.

I combined Interpretative Phenomenological Analysis (IPA) (Smith 1996, Smith, Flowers & Larkin 2022), an in-depth qualitative research approach focused on understanding how people make sense of major life experiences, with acoustic analysis of Carrie's voice to consider the following question in Part III:

- How does Carrie understand her own experiences with her voice and use her voice to navigate different interactional contexts?

To investigate this question, I conducted a semi-structured interview with Carrie, analysed with IPA, and drew on the acoustic methods developed in earlier stages of the research to compare Carrie's voice between two 20-minute conversations with different interlocutors, a close friend and an unknown interlocutor.

Overall, this research aims to demonstrate how combining auditory-perceptual analysis and multi-measure acoustic analysis, as well as quantitative and qualitative research methods, can give insight into the relationships between emergence of meaning at the micro- and macro-level variation.

Chapter 1

Outline of thesis

In the rest of Part I, I cover background to the research that is relevant to all later parts of the thesis. In Chapter 2, I explore how voice quality and phonation can be defined, set out how it is operationalised in the present study, and explore how we can describe laryngeal voice quality in terms of different phonation types. In Chapter 3, I consider how voice quality can take on different kinds of semiotic meaning (Peirce 1865, Silverstein 2003). In Chapter 4, I review auditory-perceptual and acoustic approaches to analysing voice quality.

In Part II, I use data from SCOSYA to consider how voice quality varies according to age, gender and area in speakers of Glasgow, Lothian and Insular Scots. In Chapter 6, I use Phonation Profile Analysis (PPA), a novel VPA-inspired method for analysing laryngeal voice quality that applies descriptive labels for phonation to stretches of voiced sounds, to conduct an auditory-perceptual analysis of 24 speakers from SCOSYA. In Chapter 7, I investigate how creak coded using PPA relates to creak coded automatically using an f0-based method. In Chapter 8, I investigate the relationship between acoustic measures and auditory quality in non-creaky voicing. In Chapter 9, I scale up the acoustic methods to a sample of 95 speakers to investigate variation according to social and linguistic factors.

In Part III, I conduct an in-depth qualitative case study of Carrie's voice and how she understands it. Chapter 10 gives some additional background to this part, considering why a case study of a trans speaker's voice might be valuable for understanding social meaning in voice quality, introducing Interpretative Phenomenological Analysis, and giving relevant background on trans people's experiences in the UK. In Chapter 11, I detail the methods of this section, covering recruitment via a survey, the IPA interview process and analysis, and analysis of Carrie's voice in two recorded conversations using acoustic methods. Chapter 12 gives the results of the IPA and the acoustic analysis of Carrie's voice, along with an integrated qualitative acoustic analysis of some of Carrie's 'voices' used in different contexts and how she describes them in her own

words. Chapter 13 discusses these findings, as well as proposing potential future uses of IPA in sociophonetic research.

Part IV brings together Part II and Part III with some concluding remarks.

Chapter 2

What is voice quality?

2.1 Defining voice quality

Even researchers who differ in how they use and define the terms 'voice' and 'voice quality', such as Kreiman, Gerratt & Vanlancker-Sidtis (2003: 115) and Esling et al. (2019: 1-2), agree that potential definitions exist on a continuum between 'broad' and 'narrow' definitions. Defined broadly, the term 'voice' is essentially synonymous with speech, encompassing articulation, phonation, loudness, intonation and temporal patterning, with 'voice quality' being the resulting perceived quality (Kreiman, Gerratt & Vanlancker-Sidtis 2003: 115). In the narrow definitions, such as in the technical usage by Abercrombie (1967: 26), 'voice' refers only to the 'buzzing noise' produced by the process of *phonation*: the process of vocal fold vibration that involves the vocal folds flapping open and shut. Phonation is also defined by Catford (1977: 93) as 'any laryngeal activity that has neither initiatory not articulatory function', allowing it to be extended to the laryngeal vibrations that originate from aryepiglottic and ventricular folds in the epilarynx.

In this chapter, I review ways of defining and operationalising 'voice quality' that exist along this continuum and outline my usage of terminology in the present research. Rather than viewing any definition as more or less 'correct', I take the view that how terminology is defined relates to the methods and focus of a piece of research, and that different perspectives on what voice quality 'is' must be brought together in the design and interpretation of the present research.

2.1.1 Voice quality as a quasi-permanent quality

First, consider Abercrombie's (1967: 91) definition of voice quality:

> The term 'voice quality' refers to those characteristics which are present more or less all the time that a person is talking: it is a quasi-permanent quality running through all the sound that issues from his [*sic*] mouth. These characteristics do, naturally, include some that have their origin in the anatomy of the larynx and are therefore concerned with phonation, but [...] they also include many other characteristics which have their origin elsewhere.

Abercrombie (1967: 91) defines voice quality broadly, but not so broadly that voice quality is taken as nearly synonymous with speech. For Abercrombie (1967: 91), voice quality is defined alongside two other components of speech: segmental features (the production of vowels and consonants) and voice dynamics (features such as loudness, tempo, rhythm, and pitch fluctuation). For Abercrombie (1967: 89), voice quality and these other strands are 'separable though closely woven together': voice quality is distinguished by being continually present, rather than confined to only a few segments. A secondary articulation becomes part of a speaker's voice quality if it is present as 'constant accompaniment' to a person's speech (Abercrombie 1967: 93). Similarly, the use of a particular phonation type becomes part of voice dynamics rather than voice quality if it is only present over short stretches of speech (Abercrombie 1967: 100).

2.1.2 Voice quality as a constellation of settings

Following Abercrombie (1967), Laver (1980: 1) conceives of voice quality in a broad sense and as a quasi-permanent feature of a speaker's voice, conceptualising voice quality as 'the characteristic auditory quality of an individual speaker's voice'. Central to Laver's (1980) framework for understanding voice quality is the idea that voice quality can be broken down into different componential settings. The overall voice quality of a given speaker is then 'characterized by a constellation of co-occurent settings' (Laver 1994: 402). This framework forms the basis of Vocal Profile Analysis (VPA) (Laver 1991[1979]), an auditory-perceptual method of analysing voice quality that Laver and colleagues continued to develop in later work (Laver et al. 1991[1981], Laver 1991, 1994).

In developing this descriptive framework, Laver (1980: 12) builds on the concept of articulatory setting, drawing on Honikman (1964). Honikman (1964) emphasises how particular articulatory adjustments (e.g. spread lips, open jaw) can be part of the 'articulatory setting' that creates the distinctive 'timbre' of a particular language (Honikman 1964: 73). She emphasises that articulatory setting is not just the articulation of individual sounds, but the product of their relationships and commonalities. As Laver (1978) explores, this idea dates back to early writers on phonetics in the 17th century who noted a tendencies for specific languages to be pronounced with particular

adjustments to the vocal tract. More recent work presents articulatory evidence for the idea of language-specific articulatory settings. For example, Gick et al. (2004) used X-ray to illustrate differences in inter-utterance rest position between 5 French and 5 English monolingual speakers, while Wilson & Gick (2014) found that French-English bilinguals use different inter-utterance lip, jaw and tongue positions in each language in a study with 8 speakers using optical tracking of the lips and jaw and ultrasound of the tongue.

In Laver's (1980) framework for understanding voice quality, voice quality consists of laryngeal and supra-laryngeal components, which combine to produce overall quality. The settings are compared to the 'neutral' setting, a standard reference setting that exists not as a 'normal' setting, but as a baseline against which other settings are compared. For example, the 'neutral' reference setting for phonation is modal voice: Efficient, periodic vibration of the vocal folds, without audible friction (Laver 1980: 111). Crucially, the reference setting of modal voice is not 'normal' for most speakers; Laver (1994: 414) refers to a study with 200 participants that found more than 95% of speakers used at least a degree of audible whisperiness.

Like Abercrombie (1967), Laver (1980) recognises organic (i.e. anatomical) features, like the size and shape of the laryngeal and supra-laryngeal vocal tract, and adjustments made by the speaker, as two distinct constituents of voice quality. However, Laver (1980: 10) emphasises that his descriptive system of voice quality focuses aspects that the speaker can control by adjusting the muscular apparatus of the vocal tract. He notes, for example, that his system 'largely excludes consideration of the [...] organic type of influence of voice quality', and that his model 'refers to settings of an idealized vocal apparatus, and ignores inter-speaker differences of anatomy' (Laver 1980: 7). However, organic influences can be considered through the principle of auditory equivalence (Beck 1988: 142-143), later termed configurational equivalence (Laver 1994: 426). This principle states that a similar perceptual effect might be produced by either a speaker-controlled adjustment or an organic influence such as a vocal fold disorder, but that regardless of the origin, both can be described 'by pretending that every speaker is organically the same' (Laver 1994: 427). Voice quality is then described in terms of adjustments that a speaker with a hypothetical 'standard' vocal tract would make to produce the quality (Laver 1994: 427).

Laver (1994) also departs from Abercrombie (1967) in how he conceptualises the time span that voice quality occupies. Firstly, Laver (1980: 20-22) explicitly includes dynamic and intermittent features as part of voice quality. Laver's (1994: 399) model specifically discusses how voice quality setting can be described in terms of the range and variability of a feature. Laver (1980: 22) notes that voice quality settings can be dynamic in a way that forms part of a speaker's overall voice quality, giving the example of speakers who use intermittent creaky voice when fundamental frequency

drops.

Furthermore, Laver (1980: 21) emphasises that voice quality settings are audible only on an intermittent basis and vary in prominence, because not all segments are equally susceptible to the effects of each setting. In the principle of 'susceptibility', all speech sounds exist on a gradient from non-susceptible to maximally susceptible to a particular setting (Laver 1994: 153), with segments that are maximally susceptible to a given setting being known as 'key segments' (Laver 1994: 154). For example, all segments that carry voicing are susceptible to the effects of phonatory settings, such as whispery voice (Laver 1994: 414). Furthermore, he notes that settings differ in the amount of speech data necessary for their identification: While whispery voice can be identified 'on the basis of hearing only a few syllables of phonation' (Laver 1994: 400), identification of supra-laryngeal settings usually requires more evidence.

Laver (1980: 22) also notes that voice quality can be intermittent because speakers manipulate voice quality settings for paralinguistic purposes, such as using whispery voice to signal confidentiality (Laver 1980: 22). Despite the framework's ability to account for intermittency of this kind, voice quality is still primarily a long-term, supra-segmental feature for Laver. The listener plays a key role in this conceptualisation of voice quality, with Laver (1980: 1) stating that 'perceptually, voice quality [...] is a cumulative abstraction over a period of time of a speaker-characterizing quality, which is gathered from short-term articulations used by the speaker for linguistic and paralinguistic communication. Laver (1991[1976]) emphasises that his understanding of what constitutes voice quality rather than phonetic quality is a semiotic distinction, rather than a distinction based on the underlying production mechanism or time span of the quality. For example, he notes that there is no way of telling from quality alone whether breathy voice is functioning as a phonological signal, for paralinguistic purposes such as conveying intimacy, or as a speaker-characterising voice quality (Laver 1991[1976]: 167). In this way, his framework for categorising different voice quality settings is equally applicable to short-term use of a voice quality for phonological or paralinguistic purposes (Laver 1991[1979]: 185).

2.1.3 Voice quality as the sound produced by the vocal folds

At the other end of the spectrum lie narrow usages of the terms voice and voice quality. An example of this narrow sense can be seen in the usage of voice by Garellek (2019). In that particular article, he uses 'voice' to refer only to the sound produced by the vocal folds, and 'voice quality' to refer to the resulting perceptual effects of different vocal fold configurations.

This usage is narrower than that of Laver (1980)'s not just in the sense that voice

refers only to phonation, but in the time span that it orients focus to. For Garellek (2019), voice quality is a perceptual property that is operationalised most often at the level of a single segment, usually in the study of phonemic phonation contrasts. For Laver (1980: 3), the focus tends instead to be on the overall impression of a speaker's voice that listeners gain from abstracting across variation over longer stretches of speech. In his own work, Laver (1994) distinguishes between 'phonation' and 'phonatory settings' in a comparable way. Laver (1994) includes a chapter on 'phonation', referring to 'use of the laryngeal system, with the help of an airstream provided by the respiratory system, to generate an audible source of acoustic energy' (Laver 1994: 184), which largely discusses the use of phonation to create phonological contrasts, while a later section on 'phonatory settings' in the chapter on multigsegmental settings focuses largely on speaker-characterising voice quality and language- and accent-specific settings.

2.1.4 The psychoacoustic model

The understanding of voice quality that underpins the psychoacoustic model of voice (Kreiman et al. 2014) contests the idea that voice and voice quality can be separated at all. The psychoacoustic model attempts to map the relationship between voice production, acoustic measures of voice, and listener perceptions of voice quality. Kreiman & Gerratt (2018) argue that the study of voice has been separated into voice production, voice acoustics, and voice perception, creating a situation where understandings of voice are fragmented into study at different levels of the speech chain, and so models of voice quality from each strand fail to model the voice in any meaningful way.

Firstly, Kreiman & Gerratt (2018) explore how in clinical settings, perceptual approaches to voice quality such as the Grade, Roughness, Breathiness, Asthenicity, and Strain (GRBAS) protocol tend to lack a theoretical underpinning, instead being established largely through convention. They identify Laver's model of voice quality, being based on an articulatory model of voice quality settings and their perceptual consequences, as an exception to this trend, but criticise it on the basis that 'how listeners actually use different features to assess quality, whether (or when, or why) some features might be more important than others, or how dimensions interact perceptually, is not specified' (Kreiman & Gerratt 2018: 170). More generally, they criticise perceptually-oriented models of voice quality on the basis that even expert listeners do not consistently agree in their judgements of features such as breathiness (Kreiman & Gerratt 2000, 1996) and vary their judgements based on listening context (Gerratt et al. 1993). They argue that perceptual models cannot accommodate or model these inter- and intra- listener differences (Kreiman & Gerratt 2018: 171-172).

Furthermore, Kreiman & Gerratt (2018) argue that attempts to approach voice

from acoustic or physiological approaches are often limited because they lack an underlying model to map the connection between acoustics, physiology and perception, particularly in clinical settings. They explore the example of the Acoustic Voice Quality Index (AVQI) (Maryn, De Bodt & Roy 2010), a clinical tool which brings together a number of acoustic measures to model vocal disorder severity. They outline how AVQI rests on a correlation between listener ratings of voice disorder severity and acoustic measures, but that ultimately the acoustic parameters chosen lack an underlying theoretical model to explain this correlation in terms of the connection between physiology, acoustic output, and perceived quality. Because of these issues, Kreiman & Gerratt (2018: 178) argue that it is impossible to separate voice and voice quality, preferring to characterise the voice as a singular entity so that voice exists as 'a unitary process: no voice quality, no voice production, but only voice, comprising all of these'.

The psychoacoustic model was established through several waves of research that set out to map the relationship between perception and acoustic parameters (Kreiman, Gerratt & Antoñanzas-Barroso 2007, Kreiman et al. 2014, Garellek, Ritchart & Kuang 2016, Signorello et al. 2016, Kreiman et al. 2021).

Kreiman, Gerratt & Antoñanzas-Barroso (2007) began with an assumption that the parts of the voice that are relevant to listeners would most likely be those that vary between voices and used Principal Components Analysis to identify acoustic measures of the source spectrum that varied between voices. They then refined this set of measures in a series of experiments to determine whether parameters in their model were all necessary and sufficient to model voice quality. To do this, they used analysis-by-synthesis, wherein steady-state vowels are inverse filtered, separating out the source and filter aspects of the voice, and then re-synthesized using the relevant source parameters under examination (Garellek, Ritchart & Kuang 2016, Signorello et al. 2016). This allowed Kreiman et al. (2021) to validate the model, first establishing that all parameters were sufficient by creating re-synthesized stimuli with all relevant source parameters and showing that listeners could not separate the perceptual effects of individual measures, and then establishing that all parameters were necessary by creating re-synthesized voices with fewer parameters and finding that this produced re-synthesized stimuli which were less close to original voices they were imitating.

2.1.5 Reconciling the psychoacoustic model and the componential model

The psychoacoustic model of voice exists in some ways in opposition to Laver's (1980) descriptive approach to understanding voice quality. Kreiman & Gerratt (2018: 170) argue that Laver's approach involves describing voice quality in terms of supposed, rather than actual, underlying physiological configurations of the larynx, and in terms

of where perceptual information about quality might exist based on this. However, it is worth noting at this point that the Laryngeal Articulator Model (Esling et al. 2019) (discussed further in Section 2.1.6), which draws extensively from Laver's framework, does present evidence for the physiological configurations used to produce different perceptual qualities, such as biomechanical models (Moisik & Esling 2014) and laryngoscopy (Moisik, Hejná & Esling 2019, Edmondson & Esling 2006). According to Kreiman & Gerratt (2018: 170), Laver's model of voice quality fails to consider how listeners actually assess voice quality, the perceptual importance of different aspects of voice quality, and whether dimensions interact perceptually. Furthermore, Kreiman & Gerratt (2018: 171) argue that listeners are unable to separate different dimensions of voice quality like creakiness and breathiness, a key part of Laver's model, because harmonic and inharmonic components of the voice interact perceptually (Kreiman et al. 2012).

Rather than being at odds with each other, it is perhaps more useful to see the psychoacoustic model of voice and Laver's model of voice quality as encapsulating related but separate concepts. Despite conceiving of voice quality broadly in terms of incorporating both source and filter, voice in the psychoacoustic model is narrow in terms of the time-span that it is concerned with. Specifically, the psychoacoustic model of voice was developed and validated using steady-state vowels, a limitation which occurred for pragmatic reasons to do with the inverse filtering and re-synthesis process (Kreiman et al. 2021: 464), but which represents an important difference when compared to Laver's model of voice quality. For Laver (1980: 3), a perceptual quality that exists only in a single steady-state vowel simply *isn't* voice quality, as in Laver's understanding, voice quality occurs by definition across multiple segments. The quality of a particular segment, however, compounds with the quality of other segments produced in the same utterance, or by the same speaker, or by a group of speakers, to create an overall voice quality.

Furthermore, like Laver, the psychoacoustic model of voice validated by Kreiman et al. (2021) takes a broad view of voice quality, proposing that formant frequencies and bandwidths (acoustic measures that are traditionally associated with changes in the configuration in the upper vocal tract) are necessary to model the voice fully in terms of perception. Furthermore, Kreiman et al. (2021: 464) emphasise that a strict separation of laryngeal and supra-laryngeal aspects of voice quality in terms of production is problematic, saying that speakers must adjust both the larynx and the upper vocal tract to achieve differences in voice quality or vowel quality. This is something that is emphasised by Laver (1980: 18-19), who stresses that while the settings are described separately, they interact in terms of production, perception, and acoustics.

2.1.6 The Laryngeal Articulator Model

This idea that it is impossible to fully separate laryngeal and supra-laryngeal aspects of voice quality is also at the heart of another model of voice quality, the Laryngeal Articulator Model (Esling 2005, Esling et al. 2019).

In conceptualising voice quality, Esling et al. (2019: 1) follow closely on from the understandings of voice quality put forth by Abercrombie and Laver, saying that voice quality refers to 'the long-term characteristics of a person's voice - the more or less permanent, habitually recurring, proportionally frequent characteristics of a person's speech patterns'. Esling et al. (2019: 4), however, find limitations in Laver's (1980) reliance on a linear source-filter theory for understanding voice quality. In source-filter theory (Fant 1960), vocal fold vibration creates the source of voiced sounds, and this source is independent from the filter function - the ways in which the configuration of the tongue, jaw, and lips shape this sound. Esling et al. (2019) argues that the larynx and other parts of the lower vocal tract constitute a complex articulator and resonator, analogous to the tongue in the oral cavity, because of its role in the production of pharyngeal sounds. This 'laryngeal constrictor' plays a complex role in phonation and laryngeal voice quality, with the ventricular folds and aryepiglottic folds being able to produce vibration in addition to the vocal folds. The laryngeal constrictor is also important for the production of supra-laryngeal voice quality: The actions of the laryngeal constrictor are key for tongue retraction and larynx height, which Laver (1980) includes as supra-laryngeal voice quality settings because they involve the resonating cavities above the larynx.

2.1.7 Voice quality in this thesis

In the present study, I do not adhere only to any single model of voice quality. Instead, I take the view that incorporating aspects of different models of voice quality can be helpful for understanding voice quality more fully from a sociolinguistic perspective. This idea will be explored further in Section 4, where I consider different ways of analysing and measuring voice quality, which each link back to underlying ideologies about what voice quality is.

Here, I conceive of voice quality broadly, encompassing both laryngeal and supra-laryngeal aspects and recognise that these cannot be clearly separated because of the complex role of the laryngeal constrictor. The models of voice quality discussed so far attempt to account for all aspects of what could be considered as voice quality, both laryngeal and supra-laryngeal. However, I limit myself here mainly to the laryngeal aspects of voice quality, known as *phonation*. I also use the term *laryngeal voice quality* on occasion to refer to this same concept, to emphasise that my understanding of laryn-

geal voice quality incorporates voice quality settings that are not always considered as part of phonation, including whisper, which does not involve vocal fold vibration, and harsh voice qualities, which may involve vibration of the ventricular or aryepiglottic folds. I also use this term to emphasise that I consider laryngeal voice quality as part of overall voice quality, and recognise that it is not always possible to separate laryngeal and supra-laryngeal components in terms of production, perception or acoustics. I also make use of the term 'phonation type' to refer to distinct laryngeal configurations and their resulting qualities, which can be present either at the level of the segment, or more persistently, as a phonatory setting.

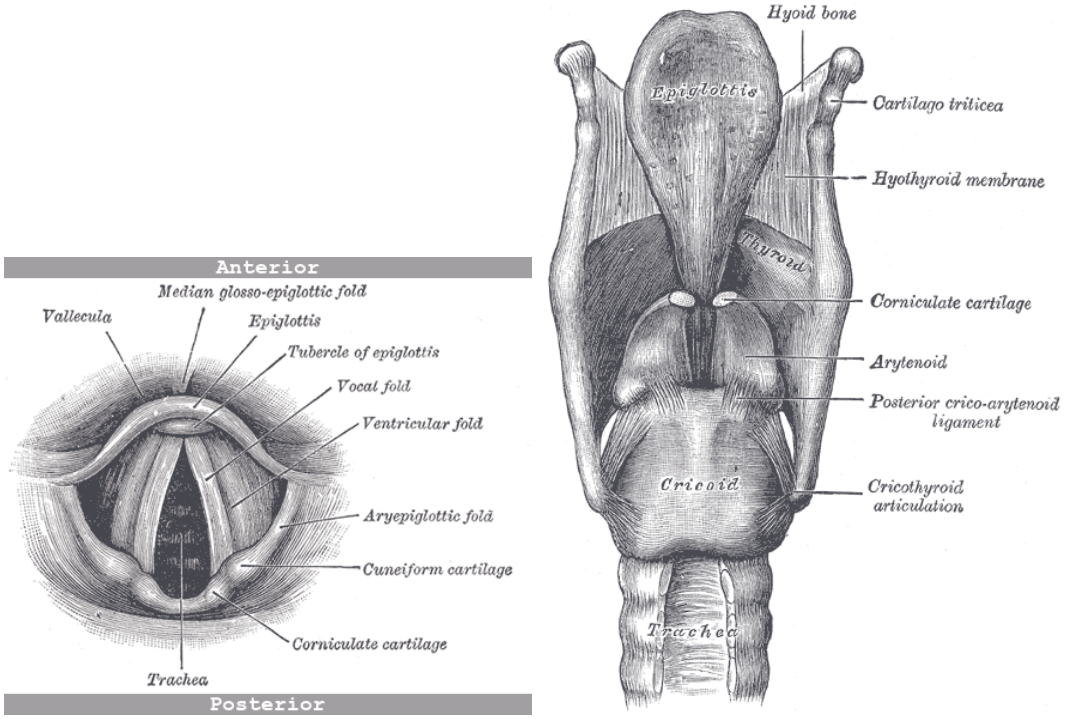
I follow Laver (1991[1976], 1994) in conceiving of the distinction between long-term speaker-characterising voice quality and short-term linguistic use of a phonation type as being one of semiotics, rather than beginning from an assumption that it is possible to differentiate between the two on the basis of quality itself. I take the time-varying aspects of voice quality as integral to voice quality itself, considering that variability in aspects of voice quality within a speaker might be a product of underlying voice quality settings in the vocal tract as well as a key characteristic of a speaker's overall voice quality. Because of this, I consider aspects of speech that are sometimes considered as part of voice dynamics, like shifts in voice quality and pitch within an utterance, as part of voice quality.

I emphasise that I also consider pitch to be part of laryngeal voice quality because of the close relationship between f0 and phonation type. Certain voice qualities settings tend to occur in certain fundamental frequency ranges (e.g. low f0 for creaky voice, high f0 for falsetto), but even where f0 is not a characteristic feature of a setting, the muscles used to alter phonation type and f0 interrelated (Zhang 2016b), leading to a close relationship between them (e.g. Kuang 2017).

2.2 Describing voice quality

Having outlined my understanding of voice quality in this thesis, here I introduce how we can describe the voice in terms of production, auditory qualities, and acoustic correlates.

Figure 2.1 shows two diagrams of the larynx which may be useful reference in the following sections. Figure 2.1b shows the structure of the larynx, including relevant ligaments and cartilages. Figure 2.1a shows a diagram of a laryngoscopic view of the larynx, showing the vocal folds as well as the ventricular folds and aryepiglottic folds in the epilarynx (situated above the vocal folds). The glottis itself, the space in between the vocal folds, is not annotated on the figure.



(a) Fig. 956 from Gray (1918: 1081): 'Laryngoscopic view of interior of larynx' (b) Fig. 952 from Gray (1918: 1081): 'The ligaments of the larynx'

Figure 2.1: Diagrams showing the anatomy of the larynx

Due to a focus on auditory quality and acoustic measurement in this thesis, I do not go into great detail on the physiology underlying these different qualities, but I refer any interested readers to Esling et al. (2019), which contains supplementary material with laryngoscopic videos of laryngeal configurations used to produce different phonation types.

2.2.1 The production of voicing

When the vocal folds are set into motion and begin to produce vibration, this is known as 'voicing'. The most widely-accepted theory of voice production is known as the myoelastic-aerodynamic theory, outlined by van den Berg (1958). The myoelastic component of this theory was first established by M'uller (1837), who conducted a series of experiments with excised human larynges and established that 'the essential cause of the voice lies in the glottis and its immediate limitation by the lower vocal cords' (M'uller 1837: 181). He attributed their ability to vibrate to their elastic nature and tension created by the movement of the thyroid, cricoid and crico-arytenoid cartilages and connected increase in the longitudinal tension of the vocal folds to rising pitch. His observations also included early remarks on the role of vocal fold tension on voice quality. He observed that where longitudinal tension is held constant, glottal aperture has no effect on pitch, but that increasing glottal aperture lead to a tone that is 'less sonorous in that one also hears the sound of the air flowing through' (M'uller 1837:

187). He noted that when medial tension is greater, 'the sound is stronger and fuller', while when this tension is lesser, 'it is weaker and more muffled' (M'uller 1837: 188).

The aerodynamic aspect of the myoelastic-aerodynamic model refers to the role of airflow in the vibration of the vocal folds, with van den Berg (1958) stressing the role of the behaviour of airflow in closing of the glottis. Laver (1980: 96) and Esling et al. (2019: 46) have summarised this process: As the vocal folds are adducted to prepare for voicing and airflow is initiated, this increases sub-glottal air pressure, overcoming the myoelastic tension adducting the vocal folds, and forces them apart, forming a narrow constriction. According to the Bernoulli effect, this narrow constriction accelerates the airflow and creates a local drop in air pressure below the folds, sucking them together, allowing the cycle to begin again.

2.2.2 Modal voice

If the arytenoids are adducted and the process of voicing is set into motion, this theoretically brings the vocal folds together along their length and creates full closure of the glottis during a cycle of voicing. If this results in periodic vibration and the laryngeal constrictor is open, and there is no additional aperiodic noise, this is known as 'modal voice' (Esling et al. 2019: 44-46).

The term 'modal voice' was popularised by Hollien (1974: 126), following communication with van den Berg, to avoid implying abnormality of other qualities through use of the term 'normal'. Modal voice is named on the basis that it include the range of fundamental frequencies that are normally used in speaking and singing: the mode. This change in terminology from 'normal' to 'modal' is apt when we consider sociophonetic studies of voice quality which demonstrate the rarity of modal voice: For example, Stuart-Smith (1999b) finds voice quality in Glasgow is characterised by a tense, whispery quality.

The acoustic quality of modal voice, and other qualities, will be explored further in Section 4.2. However, I will give a brief introduction to how acoustic measures of voice quality operate here, to allow for results of sociophonetic studies of voice quality to be described in terms of their acoustic effects in Section 3. Two main categories of acoustic measures of voice exist. Noise measures, like Harmonics-to-Noise Ratios (HNRs), jitter, shimmer, and Cepstral Peak Prominence (CPP), measure the amount of aperiodic noise in the signal. In HNRs and CPP, this is a ratio of the amount of periodic to aperiodic noise, so that highly aperiodic sounds (e.g. a voiceless fricative) show low values, while a sound that is mostly composed of harmonic energy (e.g. a pure tone, or a vowel produced in modal voice) show high values. Spectral slope measures, such as H1-H2, H2-H4, H4-2kHz, and 2kHz-5kHz, measure the drop-off in harmonic

energy at different parts of the spectrum as frequency increases, and relate to the degree of glottal constriction.

In modal voice, the amount of noise should be low, which manifests in higher numbers in CPP and HNR measures. The spectral slope is estimated to drop off at between -10 and -12 dB an octave (Flanagan 1958, Stevens & House 1961)

2.2.3 Whisper

In whisper, no voicing occurs. Instead of being fully adducted, in whisper there is a triangular opening in the glottis that comprises about a third of its length (Pressman 1942), creating a Y-shape (Esling et al. 2019: 54). Esling et al. (2019: 54) relate this Y-shaped opening to the narrowing of the epilaryngeal tube above the glottis, which shortens and narrows in whisper, narrowing the anterior glottis. The turbulent airflow generated when air flows through this narrow passage between the glottis and aryepiglottic folds (Esling et al. 2019: 53) generates the characteristic 'hissing' quality of whisper (Laver 1994: 190).

Acoustically, whisper manifests with a high amount of aperiodic noise visible in the waveform. HNR and CPP measures show very low values, as the sound generated in whisper involve noisy, turbulent airflow and lacks periodic vibration. Spectral tilt measures, which operate on the presumption of a harmonic structure resulting from vibrating vocal folds, are not valid for whisper.

2.2.4 Whispery voice

As Esling et al. (2019: 58) explain, in whispery voice, voicing is combined with aryepiglottic constriction, the central mechanism of whispery voice. The glottis is open posteriorly, but the vocal folds are adducted and set into vibration anteriorly. This glottal configuration is combined with turbulent airflow through the epilaryngeal tube. Catford (1977: 99) describes this quality as containing 'the periodic vibrations of voice at the same time as the continuous 'hushing' sound of whisper'.

The turbulent airflow and resulting noise results in lower values for noise measures than in modal voice, but higher than those of voiceless whisper. The open configuration of the glottis results in a steeper spectral slope than in modal voice.

2.2.5 Breathy voice

If voicing is initiated, but the vocal folds are abducted and the epilarynx is unconstricted, this produces breathy voice. Breathy voice involves higher airflow through the glottis, with the vocal folds vibrating relatively loosely and separated, meeting only along about half their length (Esling et al. 2019: 56). Laver (1980: 132) describes inefficient vocal fold vibration and a slight degree of audible friction. Catford (1977: 99) describes the vocal folds in breathy voice as 'flapping in the breeze', producing 'a sigh-like mixture of breath and voice'.

Laver (1980: 133) notes there is a close auditory relationship between breathy voice and whispery voice, as both involve audible friction, but the physiological mechanisms underlying them are distinct. This is illustrated by Moisik, Hejná & Esling (2019) in Figure 2.2, who present laryngoscopic images of the larynx that show a lack of laryngeal constriction and a more open, V-shaped glottis in breathy voice, contrasted with laryngeal constriction and a narrower glottis in whispery voice. Furthermore, Moisik, Hejná & Esling (2019) present laryngeal ultrasound evidence that the laryngeal constriction of whispery voice means that it is produced with a raised larynx and breathy voice with a lowered larynx.

Like whispery voice, breathy voice is also characterised by steep spectral slope and high noise. Differences between the two may involve whispery noise being more strongly characterised by noise measures rather than spectral tilt measures (Tian & Kuang 2021), steeper spectral tilt in whispery voice (Gobl 1989), and flattening of spectral tilt in higher-frequency regions due to the presence of aperiodic noise (Gobl & Chasaide 1992).

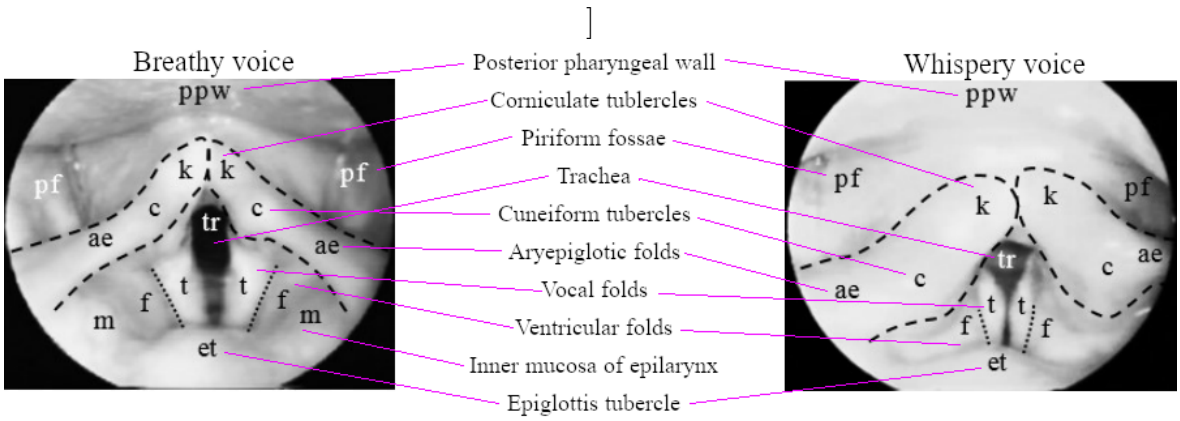


Figure 2.2: Adapted from Figure 1 from Moisik, Hejná & Esling (2019: 3). 'Laryngoscopic view of breathiness vs whisperiness'.

2.2.6 Falsetto

Auditorily, falsetto is characterised by high pitch. It is produced with the vocal folds stretched longitudinally so that they are thin at the edges, resulting in only a thin portion vibrating while the rest of the folds remains immobile (Esling et al. 2019: 61). As Laver (1980: 118-119) describes, the vocal folds often remain slightly apart, adding a whispery quality.

2.2.7 Creak and creaky voice

Catford (1977: 98) identified low frequency vocal fold vibrations as the main characteristic of creak and described its the auditory quality as 'rather like the sound of a stick being run along a railing'. He noted that the precise mechanism that produces creak is unclear, though described the vocal folds as being in close contact by not tensed, allowing air to escape from a small anterior gap. Hollien et al. (1966: 247) detailed the underlying configuration further, describing the vocal folds as thick and compressed, and positing that they make contact with the ventricular folds above them, which are also somewhat adducted. Murry & Brown Jr (1971: 446) emphasised subglottal air pressure is also lower than in modal voice.

Laver (1980) categorised creak as distinct from the combination of creak with voicing, but noted that the physiological mechanism underlying the production of creak and creaky voice remained unclear, speculating creaky voice may involve vibration of the folds at the posterior rather than the anterior end (Laver 1980: 139). Esling et al. (2019: 64) expand on the differences between creak and creaky voice and their treatment of them as a single category:

> The literature identifies a single isolated burst, or set of aperiodic bursts, as simple 'creak' phonation (Catford 1964: 32, Laver 1980: 122–126). Such occurrences are treated as voiceless, since they do not represent periodic vibration. In so far as it is possible to generate isolated bursts, the state of the larynx would be the same as for creaky voice, differing only in timing.

Here, I follow Esling et al. (2019) in grouping creak and creaky voice together and using these terms interchangeably, but recognise that further types of creak can then be distinguished based on their underlying production and acoustic manifestation. Keating, Garellek & Kreiman (2015) identify prototypical creaky voice, multiply pulsed voice, aperiodic voice, nonconstricted creak and tense/pressed voice, and describes the acoustic correlates of each type. While acoustic properties of different types of creak share characteristics, as shown in Table 2.3, there is no single acoustic property that unifies every type of creak.

Property >	low F0	irreg F0	glottal constr	damped pulses	sub-harms
Main correlate > Type ∨	low F0	high noise	low H1-H2	low noise; narrow BWs	high SHR
proto-typical	√	√	√		
vocal fry	√		√	√	
multiply pulsed		√	√		√
aperiodic	NO	√	√		
noncon-stricted	√	√	NO		
tense	NO		√		

Figure 2.3: Types of creak identified by Keating, Garellek & Kreiman (2015) and their acoustic correlates.'Table 1: Properties characterizing different kinds of creak. Check mark means a property characterizes a type; NO means it does not; blank means variable or unknown.' (Keating, Garellek & Kreiman 2015: 3). Arrows in the header point to the row or column relevant to each heading.

Keating, Garellek & Kreiman (2015) firstly outline 'prototypical creak', produced with a constricted glottis and small glottal opening, characterised by low and irregular f0 - essentially the type of creak that has been described in this section thus far. Keating, Garellek & Kreiman (2015) contrast this with 'vocal fry', sometimes used interchangeably as a term with creaky voice, which they distinguish from prototypical creak on the basis of regular f0 and damped glottal pulses. In multiply pulsed creak, there are two simultaneous f0s, one quite low and another usually around an octave above it, resulting in the auditory quality of roughness and an indeterminate pitch. Nonconstricted creak, or lax creak (Slifka 2006), involves higher glottal airflow and a more spread glottis than prototypical creak. Finally, in tense or pressed voice, the glottis is constricted but there is no drop in fundamental frequency, and no irregularity. This is consistent with what Catford (1977: 103) terms 'anterior voice', which has an auditorily 'tight' or 'hard' quality.

In this thesis, I generally treat all types of creak that are characterised by low and/or irregular f0 into one category, and distinguish this from tense voice. However, I do identify different types of creak on occasion, for example, where different types of creak lead to issues in measurement.

2.2.8 Harsh voice qualities

According to Laver (1980: 127), the label for 'harsh voice' is well-chosen, as it serves as an accurate descriptor of its auditory quality, which involves high airflow and irregular vibration (Esling et al. 2019: 69). In terms of articulation, harsh voice is characterised by epilaryngeal constriction (Esling et al. 2019: 68).

Esling et al. (2019) distinguish between several phonation types characterised by laryngeal constriction which I group here under the general label of 'harsh voice qualities'. Harsh voice, sometimes termed 'harsh voice at mid pitch', involves 'tightening the aryepiglottic constrictor mechanism, especially characterized by lower epilaryngeal tightening' (Esling et al. 2019: 67). The narrowing of the epilaryngeal constrictor results in vibration in epilaryngeal structures, and similar to whispery voice, there is the addition of turbulent noise as air is forced through a narrow epilaryngeal channel (Esling et al. 2019: 67). Harsh voice can indeed be combined with a degree of glottal opening to create harsh whispery voice. Furthermore, it is possible to combine it with creaky voice or falsetto.

In harsh voice, both the vocal folds and the ventricular folds vibrate. However, it is also possible to have ventricular fold vibration in the absence of vocal fold vibration, creating ventricular voice (Esling et al. 2019: 71). The ventricular folds are also not the only epilaryngeal structures that can vibrate: in aryepiglottic trilling, the aryepiglottic folds at the top of the epilaryngeal tube are set into vibration (Esling et al. 2019: 74). Esling et al. (2019: 75-77) also describe 'harsh voice at high pitch', a distinct quality produced when the laryngeal constrictor is engaged at the same time as pitch is raised.

2.2.9 Conclusion

Now that I have defined how voice quality will be operationalised in this thesis and outlined how we can describe laryngeal voice quality in terms of phonation types, I turn to a discussion of how voice quality can take on meaning.

Chapter 3

Meanings of voice quality

Having outlined how I conceptualise voice quality in the present study, I turn now to the other focus of this thesis: social meaning. Laver (1991[1976]) observed that there is no *a priori* way to distinguish between linguistic, paralinguistic and extralinguistic meaning in laryngeal voice quality. I therefore consider here how voice quality takes on different kinds of meaning, how we can study this, and how both macro-level and micro-level approaches can bring value to the study of social meaning in Scottish accents in the present study.

In theorising how voice quality takes on meaning, many authors draw on Peirce's theory of semiotics (Abercrombie 1967, Laver 1991[1976], Silverstein 1976, 2003). In this framework, signs can be differentiated in terms of their relationship to what they signify, allowing us to explore how depending on context, phonation can be taken to signal linguistic meaning, speaker traits or group members, or resemblance to other sounds.

Peirce (1865) describes three types of sign. Symbols bear an arbitrary relationship to the entity that they signal: For example, the English word 'fire' is used to represent the concept of fire by convention. Phonation takes on *symbolic* meaning when it distinguishes the words /bar/ ('twelve'), and /b̤ar/ ('outside') in Gujurati (Ladefoged 1971).

An index signals something about the context it usually occurs in. As we often see smoke in the presence of fire, we take smoke to *index* a nearby fire. Where male speakers from the North of England use creaky voice extensively (Henton & Bladon 1988), listeners may then interpret their use of creak as an *index* of their group membership of Northern male speakers.

Meanwhile, 'likenesses', or *icons*, have a 'community in some quality' to the thing they represent (Peirce 1865: 294). The icon displayed in Figure 3.1 is not, in and of

Figure 3.1: An iconic representation of a fire.

itself, a fire, but bears some resemblance to it in colour and form. Phonation is can take on *iconic* meaning if we creak to imitate the sound of a frog.

Laver (1991[1976]) drew on Peirce's semiotics to explore how the total sound a speaker produces includes *intrinsic* features, which derive from the physical characteristics of a speaker's vocal tact, and potentially controllable *extrinsic* features (Laver 1991[1976]: 167), together forming a *concurrent* background which *exponent* features can then occupy to take on phonological or paraphonological meaning. The decision about labelling a particular feature as part of one of these categories rather than an other, he argues, is 'necessarily a semiotic decision about function rather than a descriptive decision about substance' (Laver 1991[1976]: 167). The ability for an equivalent voice quality feature to be understood as part of a speaker's intrinsic or extrinsic voice quality or as communicating phonological or paraphonological meaning, makes it an example of what Silverstein (1976) terms a *shifter*. Silverstein (1976) also drew on Peirce's theory of semiotics in his exploration of language as a system of signs, terming features for which the referent shifts depending on factors of the speech situation 'shifters' (Silverstein 1976: 24). The meaning that a shifter takes on, Silverstein (1976) argued, is constituted by the speech contexts that it occurs in.

Here, I explore how phonation can be used to convey symbolic, indexical and iconic meaning, dependent on its linguistic and interactional context, and focus specifically on aspects of phonation that convey social meaning to consider the value that different approaches to the study of social meaning bring to the present study.

3.1 Symbolic: Segmental functions of phonation

Symbolic meaning is conveyed through phonation in many languages through phonemic phonation contrasts. Many linguistic contrasts in terms of phonation can be described in terms of the oppositions between voiced-voiceless, and aspirated-unaspirated (Ladefoged 1971: 9), as in the contrast between English voiceless [p, t, k, f, s, ʃ] and voiced [b, d, g, v, z, ʒ] (Catford 1977: 95-97). In plosives, this contrast is one of aspiration

following release, rather than voicing during closure: English [p^h,t^h,k^h] involve a period of voicelessness during the release and immediately after the release of a glottal stricture (Ladefoged 1971: 9), while voicing begins more quickly after release of closure in unaspirated English [b, d, g].

In a survey of 567 of the world's languages, Maddieson (2013) finds that 32% of languages have no voicing contrast in either vowels or fricatives, compared to 33% which have a voicing contrast in plosives but not in fricatives, 28% in both plosives and fricatives, and 7% in fricatives but not plosives. English falls into the 28% that has a contrast in both plosives and fricatives. Wells (1982: 409) and Stuart-Smith (1999b: 208) note that in Scottish accents, the subject of this thesis, voiceless plosives are less aspirated: That is, the portion of breath that follows the release of a voiceless plosive before the onset of voicing in a following vowel is shorter than it is in many other accents of English.

However, in Gujurati, Danish, Yi, and White Hmong, and many other languages, non-modal phonation is also used to produce phonemic distinctions (Gordon & Ladefoged 2001, Esposito & Khan 2020). Ladefoged (1971) conceptualises phonemic phonation contrasts as occurring along a one-dimensional model ranging from full glottal closure (glottal stop) to voicelessness, with other states such as creak, tense voice, lax voice and breathy voice occurring at different points along this continuum. As Esposito & Khan (2020) note, many languages distinguish between two points on this continuum, contrasting voiced and voiceless sounds, but countless more languages distinguish between further points, and some also incorporate epilaryngeal constriction into their phonemic system. Phonemic phonation contrasts can occur only on vowels (e.g. Burmese), only on consonants (e.g. Javanese), or on both vowels and consonants (e.g. White Hmong), and the precise timing, duration and acoustic and auditory quality of contrastive non-modal phonation can vary between languages (Esposito & Khan 2020). Phonation contrasts can also have relationships with vowel quality, tone and airstream mechanism (Esposito & Khan 2020).

Though English is not typically considered to have phonemic phonation contrasts beyond aspiration and voicing contrasts Paolo & Faber (1990) explored a potential example in Utah English in an ongoing tense-lax vowel merger before /l/ (e.g. full-fool). They presented perception and production evidence that as differences in vowel quality lessened, speakers may have instead produced differences in phonation that listeners drew on, alongside formant frequencies and vowel duration, as a cue to vowel quality.

Non-modal phonation can also take on symbolic meaning through allophonic variation. While Catford (1977: 98) notes that glottal stops are not part of phonation as they are involved in articulation or initiation, they can be realised as a portions

of creaky voicing or favour creak in surrounding contexts. Ladefoged & Maddieson (1996: 75) presented an example from Lebanese Arabic where a speaker used creaky voice in place of complete glottal closure, and stated that this phenomenon occurred 'in the great majority of languages [they] have heard'. Docherty & Foulkes (1999) also noted this phenomenon in Newcastle English, and found that only 3% of glottal variants (which also include cases of the Tyneside glottalised variants) featured a canonical voiceless stop closure, and instead were characterised by laryngealisation and acoustic cues to creak. This may be relevant in the present work, as glottal stops are notably common in the Scots varieties studied here (Stuart-Smith 1999a, Schleef 2013, McCarthy & Stuart-Smith 2013, Sundkvist 2011, Schmitt 2015).

Breathy voice can also occur allophonically in English. Roach (2004) reports that in Received Pronunciation (RP), /h/ is produced as the voiced glottal fricative [ɦ] in a voiced environment (e.g. *behind*). Ladefoged (2006: 143) notes this phenomenon in English more generally too, and argues that this voiced glottal fricative can be best understood as a period of breathy voice or murmur. /h/ can also favour the occurence of breathy voice in adjacent vowels through co-articulation. Epstein (2000) found breathiness occurring in 4 male and 4 female English speakers after /h/. Garellek (2012) argues that co-articulatory effects of this kind in English are distinct from 'true' phonation contrasts: They compare the effects of aspiration and pre-glottalisation in English to contrastive non-modal phonation in Hmong, and find that non-modal phonation in English arising from coarticulation is present during less of the vowel that contrastive non-modal phonation in Hmong.

In Scottish accents, a word-final voiceless fricative in Scottish accents can also favour pre-aspiration (Gordeeva & Scobbie 2010), which can appear as a period of breathier voicing or voicelessness towards the end of the vowel, related to the vocal folds opening in anticipation of a voiceless segment (Ladefoged & Maddieson 1996).

I will return to allophonic and coarticulatory effects in more detail in Chapter 5, where I formulate predictions for linguistic effects in the study. I turn now consider how voice quality takes on indexical meaning of different types, such as a speaker's physical characteristics; membership of a geographic region or accent group; membership of a social category such as social class, ethnicity, or gender; an evaluation of a feature of the interaction, known as stance (Bois 2007); emotional states; or a particular character, voice, or figure that a speaker aligns themselves with in a particular interaction, known as a persona (Agha 2005).

3.2 Indexical: Social meanings of voice quality

Abercrombie (1967: 6-9) divided indexical markers in speech into three classes:

(a) Markers that indicate group membership, including *regional* and *status indices*, as well as learned components of gender variation. This can extend to any social grouping that uses a particular feature, such as occupation.

(b) Markers that characterise the individual, arise from vocal tract physiology (e.g. sexual dimorphism)

(c) Those that reveal changing states of the speaker, such as the speaker's temporary emotional or physical state

Abercrombie (1967: 92) distinguished speaker-controlled components of voice quality, originating from muscular tension settings, and those outwith a speaker's control. Abercrombie (1967: 92) proposed that voice quality components that are outside of a speaker's control cannot be indexical of social groups, but rather of speakers that share physical characteristics (e.g, a particular quality indexing that someone has laryngitis). However, Abercrombie (1967: 94) noted that in practice learned and unlearned components may be intertwined, suggesting that the velarised voice quality characteristic of Liverpool (Knowles 1973: 89) may be learned as a result of a historic prevalence of adenoids, a physical cause of adenoidal voice quality that can be imitated using velarisation.

Laver & Trudgill (1991[1979]) draw on Abercrombie (1967) in describing how speech can act as a social marker while being a vehicle for linguistic communication. They propose the following distinctions in addition to Abercrombie's typology:

(a) *Social markers*, that index social characteristics such as 'regional affiliation social status, educational status, occupation and social role'

(b) *Physical markers*, that index 'physical characteristics, such as age, sex, physique and state of health'

(c) *Psychological markers*, that index 'psychological characteristics of personality and affective state'

Later research demonstrates that listeners do use voice quality information to make judgements of these types, with Yuasa (2010) finding creaky voice led young women to be perceived as urban-oriented and upwardly mobile, Pessin et al. (2017) finding auditory-perceptual ratings of voice problems reflected self-rated voice problems and physical differences observed with laryngoscopy, and Gobl & Chasaide (2003) finding connections between voice quality and perceived affective states.

However, research also suggests that these strands of information are highly interconnected. Comparing how 82 Irish English, Russian, Spanish, and Japanese listeners

attributed emotion to similarly synthesized Swedish stimuli, Yanushevskaya, Gobl & Chasaide (2018) found that although there was some cross-linguistic similarity in how listeners attributed emotion across languages, there were also differences, with tense voice playing less of a role for Japanese and Spanish listeners than for Irish English and Russian listeners. This suggests that any information about psychological state present in the voice is necessarily intertwined with social information.

Furthermore, differentiating between these strands of information is not a straightforward task for listeners. Laver (1991[1968]: 156) discusses the role of voice quality in judgements listeners make about speakers on depends on their prior experience and cultural background, and emphasises that there is no guarantee that judgements are a reliable indication of a speaker's characteristics. Listeners may, for example, attribute social characteristics to speakers on the basis of vocal characteristics that have a biological basis, and vice-versa. For example, Matar et al. (2016) considered how listeners perceived the gender of women with Reinke's Edema, a vocal fold disorder characterised in part by a change in voice quality and lower pitch, who often report being perceived as male over the phone. They found that listeners were more likely to perceive female speakers with Reinke's Edema as 'surely masculine' than 'surely feminine', compared to women without Reinke's Edema.

Theories of indexicality including the indexical order (Silverstein 2003), indexical fields (Eckert 2008), and iconization (Irvine & Gal 2000), provide a framework for relating micro-level instances of hearing a speaker's voice in an interaction to the development of associations between voice quality and macro-level social categories. Here, I draw on these theories to discuss how voice quality takes on social meaning.

Silverstein (2003) proposes that we need a framework for understanding how use of sociolinguistic variables in micro-level interactions relates to macro-level social categories of identity. He proposes the concept of *indexical order* to describe how correlations between a linguistic form and a social grouping can become abstracted, be transformed, and take on new meanings. At one level of meaning (the level that Silverstein calls the n-th order) the use of a feature can be connected to some kind of semantic, stylistic, or social-demographic category that could be visible to an outsider, but is unremarked upon in-context by users because of its appropriateness to that context. Johnstone & Kiesling (2008: 8) explain the orders of indexicality with reference to the linguistic situation in Pittsburgh, giving the example of a speaker using variants that can be correlated with macro-social categories, such as being a working-class male speaker from Pittsburgh, which they term 'first-order' indexicality. These correlations are noticeable to sociolinguists as outsiders, but unremarked upon by speakers with dense social networks. This meaning can then be filtered through context, the other linguistic forms it appears with, and ideology, to take on new, transformative meanings. At this next level of indexical meaning (which Silverstein refers to as n + 1st

order), a feature is assigned some kind of meaning that speakers themselves are aware of, mediated by ideology; In Johnstone & Kiesling's (2008) Pittsburgh example, they identify 'second-order' meaning when speakers notice their own use of regional features and can shift them between different contexts. Each order of indexicality can be added to, and another level of meaning can always theoretically exist when features at one order come to be seen as meaningful in terms of another ideology. This underlies Silverstein's usage of the ordinal n to present the concept of indexical order, rather than simply referring to first and second order: Any level of indexical order can be built on, becoming the n-th order and allowing an n + 1st order meaning to emerge. Johnstone & Kiesling (2008) give the example of a third-order indexicality emerging in Pittsburgh when people begin to notice second-order variation and indexical meaning and remark on it and codify it, such as in the case of a stereotypical local Pittsburgh character created by a radio DJ who is sometimes referred to as 'Yinzer' a term derived from the local plural second-person pronoun *yinz*.

As Silverstein (2003: 227) notes, much sociolinguistic research proceeds with what he terms '1st-order' analysis. This type of analysis presupposes the existence of relatively stable macro-social categories such as class, race, and gender, and that they are useful for understanding the meaning of variation. However, Silverstein (2003: 227) emphasises that considering only a single indexical order in any analysis 'gives us no interesting insight; it is an indexical partial, a beginning, at best.'

Much of the research that Eckert (2012) describes as belonging to the First Wave of sociolinguistic variation research might be seen as taking a 1st-order approach. Eckert (2012) differentiates three waves of analytical practice in variationist sociolinguistic research, where the first wave concentrates on correlations between variation and macro-social categories, the second on locally-situated categories informed by ethnographic methods, and the third on stylistic practice and ideology. Eckert (2012: 88) notes that while Labov's (1963) research in Martha's Vineyard was 'all about social meaning', taking ethnographic observations and interviews into account, First-Wave research that followed this came to depend heavily on macro-social categories in explanations of variation. It is worth noting that First-Wave research conducted before Silverstein (2003) formally introduced the concept of indexical order often still engaged with higher-order indexical meaning and the relationships between different levels of meaning, however. Silverstein (2003: 217-220) comments on how Labov discusses indexical order through the terms 'indicator', 'marker' and 'stereotype', for example. He uses the term 'indicator' to refer to variables that characterise a speaker's membership of a macro-social category, but 'marker' to refer to variables that show within-speaker variation between different stylistic contexts, and 'stereotype' to refer to markers that have become imbued with ideology, which speakers draw on in imitation of stereotypical personae. First-Wave research is not equivalent to 1st-order indexical analysis, but criticism of both often centres on over-reliance on researcher-defined categories for

explanations of variation.

In the rest of this section, I consider how meaning can be seen at different levels of indexical order in different studies on voice quality. In doing so, I intend to explore how voice quality varies in terms of macro-social categories in different varieties, and discuss the limitations of first-order investigations, as well as the value of investigating how voice quality varies according to macro-level social categories. I turn then to discussion of how voice quality can be used and interpreted as an *icon*, the final type of Pierce's signs.

3.2.1 The first order: Voice quality and socio-demographic categories

Early sociolinguistic research on voice quality considered how articulatory settings characterised particular accents and communities. The earliest analyses considered how patterns of segmental features revealed potential underlying settings. Labov (1963) considered how 29 speakers from urban, down-island areas and 40 speakers from rural, up-island areas in Martha's Vineyard produced the first part of the diphthongs /ai/ and /au/. He found that up-island speakers produced more centralised variants. Combining this with evidence from other phonological variables, he noted a tendency for up-islanders to produce higher, more constricted variants, and down-islanders to produce lower, more-open variants. He identified economic pressures on the up-island fishing community as a contributing factor, allowing features associated with island identity to be used as to oppose the growing dominance of the tourist industry in urban areas. He suggested rather than centralised diphthongs being independently socially evaluated by listeners, 'this "closed-mouth" articulatory style is the object of social affect' (Labov 1963: 307), and that this tendency towards a certain articulatory posture may play a role in sound change.

Knowles (1973) looked at voice quality and articulatory setting in 47 Scouse speakers in Liverpool. As well as describing supra-laryngeal articulatory settings and reflecting on the 'adenoidal' quality of Scouse, Knowles (1973: 115) gives impressionistic remarks on the laryngeal voice quality of Scouse, saying 'perhaps Scouse has a whispery voice' and noting the presence of idiosyncratic use of creak by some speakers.

Trudgill (1974) also considered articulatory setting in 60 speakers from Norwich. Trudgill (1974: 186) noted a tendency for working class speakers to employ creaky voice, as well as raised larynx and high muscular tension which combine to produce a 'harsh and metallic' quality (Trudgill 1974: 187). As Esling (1978b: 19) notes, this analysis was based on an overall impression of the voices of all the participants in the sample rather than a systematic analysis of each individual's voice quality, but

presents an important preliminary application of the descriptive labels developed by Laver (1991[1968]) and Abercrombie (1967) to sociolinguistic data.

Research on Scottish voice quality, discussed in more detail in Section 5.1, developed this approach, with Esling's taking Labov's systematic approach to the study of phonological and syntactic variation and applying it to the study of voice quality using Vocal Profile Analysis (discussed as a method further in section 4.1). Esling (1978b) considered variation according to socio-economic status among boys (n=18) and fathers (n=32) from Edinburgh and found greater incidence of creaky voice in the group with the highest socio-economic status, but more harsh voice in the group with lowest socio-economic status. Esling (1978b) also considered stylistic variation between read and interview speech, finding that working-class participants do not approximate middle-class voice quality when reading. Instead, participants tend to use the same settings as they do in interview speech, but on lower scalar degrees, suggesting that speakers vary their voice when reading, but do not model features associated with higher social class backgrounds in read speech. In fact, all groups used less creaky voice (used most by the highest socio-economic status group) in read speech.

Phonetic research based outside of a Labovian-sociolinguistic approach looked to physiological rather than social explanations for variation in voice quality. This is evidenced in research on gender variation in voice quality. Monsen & Engebretson (1977) looked at differences in harmonic spectra and the shape of the glottal source waveform in 5 male and 5 female participants and found steeper drop-off of harmonics and a more symmetrical glottal wave shape, which both indicate a more open glottal configuration and breathier quality. They interpreted these findings in terms of differences in vocal fold size, arguing differences in vocal fold mass affect how the vocal folds meet during vibration and cause differences in the resulting glottal waveform (Monsen & Engebretson 1977: 992).

Henton & Bladon (1985) considered whether increased breathiness is 'normal' in female speakers by looking at male and female adult speakers in two accents, 36 speakers of RP and 25 of Modified Northern (speakers who grow up in or near Leeds but later moved away). Comparing the first and second harmonics of vowels, they found a mean difference of 5.5 dB between male and female speakers in RP, and 5.1 dB in Modified Northern speakers. This suggested female speakers used increased breathiness in both accents, and they stress the importance of taking this variation into account when establishing what is considered to be a 'normal' degree of breathiness in non-pathological speech.

Henton & Bladon (1985) suggest that a breathy voice is 'a sexy voice' and 'associated with arousal' due to physiological processes of lubrication during arousal that affect the whole body including the larynx, causing the vocal folds to vibrate ineffi-

ciently, and that women 'imitate the voice quality associated with arousal' in order to 'be regarded as more desirable or with greater approbation by a male interlocutor'. Incorporating social relationships as a driver of gender variation here departs from the primarily physiological explanation given by previous researchers such as Monsen & Engebretson (1977). However, interpretation of findings shows that relying on gender alone, in the absence of other data to help explain variation, can cause issues. Although Silverstein's (2003) framework of indexical order did not exist at the time that Henton & Bladon (1985) conducted their research, this explanation demonstrates the issues of conducting analysis only at a single level of the indexical order and attempting to explain findings with reference to these categories. Finding a gender difference in breathy voice, they turn to explaining this in terms of a meaning that exists at a different level, that of arousal, and suggest that women are aware of this, and exploit it for social desirability. This is a difficult conclusion to justify with reference to the methods employed in the study. Though there is some limited evidence for hormonal effects on female voice quality at different stages of the menstrual cycle (Hejná 2019, Raj et al. 2010), Henton & Bladon (1985) do not demonstrate or evidence this connection, and in neglecting to do so demonstrate a potential pitfall in taking socio-demographic categories as the main focus of analysis.

Henton & Bladon (1988) consider whether creak, too, varies by gender (male or female) and accent (RP or Modified Northern). They find that male speakers, in both RP and Modified Northern, use more creak, with male Modified Northern speakers using more creak than RP male speakers but female speakers of each accent using comparable amounts. Male and female RP speakers are thus less differentiated than Modified Northern speakers in terms of creak, which Henton & Bladon (1988) suggest could be related to a previous finding by Elyan et al. (1978) that RP accents are perceived as more androgynous than Northern accents. They note that in their previously discussed study on breathiness, they also found less of a gender difference in RP speakers compared to Modified Northern speakers. Again, findings in terms of gender variation are difficult to interpret. They present two possible interpretations: Either Modified Northern speakers use creak to exaggerate a gender difference, or RP speakers use it to underplay a gender difference. They suggest that more studies on baseline levels of creak would be necessary to know which interpretation is true. With the benefit of access now to many more years of research on creak, including some that find female speakers using more creak in some sociolinguistic contexts (Yuasa 2010), it seems possible that neither of these interpretations is quite 'true', and that rather, investigating why male Modified Northern speakers use so much creak would require a different approach.

Considering how voice quality varies according to socio-demographic categories can help to identify macro-level patterns. Stuart-Smith (1999a) describes characteristic features of voice quality in 32 speakers from Glasgow stratified by age, gender and social

class. Stuart-Smith (1999a) identifies a stereotypical 'Glasgow voice' used in television characters such as Rab C. Nesbitt, that involves harsh phonation. Stuart-Smith (1999a: 211) notes that this stereotypical 'Glasgow voice' is often impressionistically described as 'rough' and is associated with anger and violence. Stuart-Smith (1999a) goes on to investigate whether there is any evidence for stereotypical Glasgow voice quality. She finds little harsh voice in Glasgow, with Glasgow voice quality instead characterised by a cluster of settings, including tense, whispery phonation. Glasgow voice quality also varied by gender and class: Male speakers used more creaky voice and female speaker more whispery voice, while working class speakers also used more whispery voice. Considering variation according to these categories in Glasgow allows Stuart-Smith (1999a) to demonstrate that there is little evidence for the stereotypical 'Glasgow voice', while considering shared features of Glasgow voice quality and showing the existence of variation within a community.

The research discussed thus far appears to show a link between creaky voice and male speakers and breathy voice and/or whispery voice with female speakers. Additional evidence for this comes from Klatt & Klatt (1990) who considered whether incorporating breathy voice into speech synthesis would improve synthesis of female voices. They looked at ten female and six male speakers of American English. As well as being rated as breathier than male speakers, female speakers displayed acoustic cues to increased breathiness such as higher H1, aspiration noise in the region of F3, and wider first formant bandwidth. Similarly, Price (1989) looked at spectral and glottal source characteristics of 4 male and 4 female speakers of American English in monosyllables, they found shorter closed phases and less energy in higher-frequency parts of the spectrum in female speakers, suggesting increased breathiness.

Other research problematises this connection. Beck & Schaeffler (2015) conducted Vocal Profile Analysis for 76 male and female adolescents in Dumfries, Aberdeen and Inverness, and in line with Stuart-Smith (1999a), found male speakers using more creaky voice and harsh voice, but no difference in any other phonatory quality. They suggest instability in the larynx during puberty and emphasis of sexual dimorphism through the use of low-f0 creak may both contribute to adolescent male speaker's use of creak. This possibility is supported by similar tendencies in supra-laryngeal voice quality settings, with male speaker potentially using increased lip rounding and larynx to emphasise vocal tract length.

Syrdal (1996) looked at voice quality in 5-7 second excerpts of telephone conversations in 160 speakers of American English. They found that H2-H1 was significantly higher more male speakers (mean = 13.7 dB), than for female speakers (mean = 4.7 dB). H2-H1 can be interpreted in the opposite direction to H1-H2, suggesting that female speakers were more breathy. They also identified instances of creak through auditory and spectral cues, and found that female speakers used more instances of creak

(mean = 1.8) than male speakers did (mean = 1.2) in each each excerpt. This emphasises that being more breathy or more creaky is not a simple dichotomy; Speakers can be both *breathier* and use *more creak*.

Considering more accents and varieties of English, and incorporating race and ethnicity as a socio-demographic variable rather than concentrating on white speakers, complicates this pattern further. Szakay (2012) looked at voice quality in Maori and Pakeha (White European) New Zealand speakers, looking at 15 seconds of speech from 36 speakers. Using H1-H2, for which higher values can be interpreted as more breathy or laxer and lower values can be interpreted as tenser or creakier, she found that Maori speakers and older speakers tended to use a creakier phonation. However, older Maori male speakers did not fit this pattern: as age increased for Maori male speakers, speakers showed higher H1-H2 values, indicating breathier phonation. For Pakeha speakers, there was a gender difference regardless of age, but male speakers tended to be breathier and female speakers tended to be creakier.

In UK varieties, too, ethnicity can be a factor. Discussed in more detail in Section 4.2, Szakay & Torgersen (2015, 2019) compared voice quality between speakers from Anglo backgrounds from outer London, and speakers both Anglo backgrounds and immigrant backgrounds in inner London, and found a complex relationship between voice quality and social factors: gender, inner/outer London, and ethnic background.

Considering variation according to macro-level social categories also helps to demonstrate the variation that exists within genders across speakers of different groups. Wileman (2018) investigated voice quality among 36 young middle-class female South African English speakers in a joint auditory-perceptual and acoustic study. He finds that voice quality varies according to ethnicity on a range of spectral tilt and noise measures, with black speakers using a voice quality characterised by breathiness, compared to white speakers using a more tense or creaky quality.

Considering variation according to macro-level categories also provides a backdrop for further investigations that consider how social meaning becomes attributed to that variation and the relevance of that variation to more locally-defined categories. Podesva (2013) considered how use of voice quality varied in 32 white and Black American male and female speakers in the Washington DC metropolitan area, combined with analysis of what speakers were communicating when they used a particular quality. He found that female speakers used more creaky voice than male speakers, regardless of race, and no gender difference in terms of breathy voice. In terms of whispery voice, he found that only white men were marked in terms of whispery voice, using it less than any other group. Unlike much other research, he also considers use of falsetto, and finds that Black women use falsetto more than any other group. First identifying macro-level patterns enabled him to then analyse the stances taken when falsetto was used to gain

insight into its social meaning, discussed below in Section 3.2.2.

Looking at macro-level variation also allows investigation of social meaning on a large scale. Gittelson, Leemann & Tomaschek (2021) investigated how phonation varied in 2,500 British English speakers in read speech using participant-recorded smartphone data. They found that phonation varied according to age, sex, and education in an investigation using multiple measures of phonation: F0, HNR, H1*–H2* and CPP. They found that male speakers were creakier than female speakers, showing both lower HNR (=less periodic) and lower H1*–H2*, which together can be taken as suggesting increased creakiness. They also found that sex and age interacted: Older male speakers showed decreased HNR and H1*–H2* with age, while female speakers showed the opposite. Furthermore, speakers with a higher level of education were creakier, which Gittelson, Leemann & Tomaschek (2021) suggested may relate to more educated speakers being more likely to use their voices in their occupation (e.g. teachers, singers), potentially increasing prevalence of voice disorders.

Though I focus mostly on varieties of English in this review of the literature on indexical meanings of voice quality, cross-linguistic comparison also reveals that connections between voice quality and gender are not universal. Hejná et al. (2021) looked at voice quality in multi-measure acoustic analysis of read speech 120 male and female speakers of Czech and Danish. They found that female speakers were breathier according to CPP, H1*–H2* and H1*–A3*. Male speakers, however, showed lower HNR and higher H2*–H4* than female speakers, which would both usually also be interpreted as showing increased breathiness. This apparently contradictory finding becomes clearer in light of a perceptual task they conducted alongside this experiment, where three speech therapists rated the voices on a scale of breathiness to roughness. They found male voices were more likely to be rated as rough, while female voices were more likely to be rated as breathy, and that HNR and H2*–H4* correlated with perceived roughness. They also found differences in the degree of magnitude of difference between language when comparing their results to those of previous studies (Chen et al. 2010, Hanson & Chuang 1999, Garellek et al. 2013), supporting the idea that there is a degree of language-specificity in gender variation in voice quality.

Together, these studies demonstrate that there is no one-to-one mapping between voice quality and social meaning: Creaky voice, for example, is not consistently connected to male or female speakers. Eckert's (2008) concept of the indexical field is useful for exploring this idea in more detail. Eckert (2008) notes researchers commonly interpret findings as reflections of speakers' membership in social categories, but argues that variables have more general meanings that become more specific when they occur in context. In this way, breathy voice might have a range of indexical meanings, and the precise meaning that a listener attributes to an instance of breathy voice depends its context - what variables and variants it co-occurs with, the wider style of a speaker,

and the listener's own background and orientation to that context.

3.2.2 n + 1st order variation and approaches

Mendoza-Denton (2011) sets out to ask a question specifically concerned with the role of context in constructing the social meaning of voice quality: 'How does a variable travel from context to context and become part of the constellation of features that indexes a particular persona?'. She notes that creaky voice is associated with masculinity within phonetics, but that some research links it to female speakers, and considers how creaky voice might travel between different meanings and contexts to enable this. She argues that creaky voice takes part in a process of semiotic hitchhiking: As a supra-segmental feature, it cannot occur in isolation, and thus has no 'vehicle' of its on to circulate in, so it hitchhikes with salient linguistic and non-linguistic features that form part of a style, stance or persona. She identifies the use of creaky voice in a narrative told by a teenage Chicana girl, Babygirl, who uses creaky voice in the construction of a 'hardcore' gang persona. She argues that Babygirl uses creak to express ideologies of acting hardcore, and argues that creak she uses creak to regulate emotion rather than to index masculinity. In the years since Babygirl's narrative was recorded, though, she argues that the idea of a hardcore Chicano gangster persona circulated in rap music, the media, online and in video games. She argues this allows creaky voice to hitchhike as this persona developed, leading indirect indexicality of creak to shift from a hardcore persona on the local level in Babygirl's narrative, to a stereotypical Chicano masculinity in the popular imagination.

Podesva (2007) considers how voice quality becomes associated with different personae and identities by looking at how a single speaker uses voice quality across interactions. He takes the case of Heath, a white American medical student in his twenties, and examines how he uses his voice in a barbecue with friends, a phone call with his father, and a meeting with a patient. These situations differ in terms of formality and interlocutor, but also in terms of the persona Heath presents: While Heath's identity as a gay man is not particularly relevant in the patient meeting, and not often brought up in conversation with his father, with his friend group he presents a particular flamboyant, image-conscious, 'diva' persona. At the barbecue with his friends, Heath uses more falsetto, longer stretches of falsetto, with higher maximum f0, often accompanied by creaky voice.

Podesva argues that Heath's use of falsetto and creak are part of how he creates this persona. He considers the functions of Heath's use of falsetto, and finds that they are unified by expressiveness, as well as the topics that Heath is talking about when he uses this expressive falsetto. He finds that Heath often uses expressive falsetto often when he performs these parts of his personality – expressing excitement at his friend's

dress, or yelling when someone tries to mess up his hair. Podesva (2007) speculates that because expressing emotion is often seen as a non-normative behaviour for men, by performing expressiveness with falsetto, Heath relinquishes heteronormative masculinity and creates the opportunity for listeners to interpret a link between falsetto and gay identity.

Other work has considered how voice quality is used in the portrayal of characters in films, television shows, and table-top roleplaying games. Despite not being natural speech data, this type of analysis is useful for considering social meaning because, as Moisik (2013: 194) notes, they allow us to consider 'the subtle, and sometimes not so subtle, ways that language is used to create and disseminate sociocultural ideologies and values'. Moisik (2013) considers cases of harsh voice in characters represented as black in American media by both black and non-black performers and finds examples of harsh voice being used by performers to contrast stereotypical portrayal of black characters from those of white ones. Rather than reflecting true sociolinguistic patterns, Moisik (2013) argues that performers use harsh voice in their portrayals of black characters to convey a heightened affective state, portraying stereotypical black characters as aggressive. He also suggests that it is used in some cases to emulate the voice effects of smoking.

Boyd & Hejná (2022) consider how a single voice actor, Matthew Mercer, uses voice quality in his portrayal of 19 different characters in Critical Role, a recorded Dungeons & Dragons campaign. They find that Mercer uses voice quality to portray characters stances as good or evil, with allies to the players using more breathiness but enemies and non-human characters using more whispery phonation. Outside of English, Starr (2015) illustrates how phonation contributes to the construction of 'sweet voice' in Japanese anime, a quality often used by voice actors to portray supporting female characters in positions of feminine authority, like older sisters.

Additional value comes from combining analyses of the function and meaning of voice quality at the interactional level with larger scale analysis. As discussed in Section 3.2, Podesva (2013) conducted a quantitative analysis of variation in voice quality by race and gender in Washington DC and found that Black women used falsetto more than any other group, which then allowed him to consider the interactional function of falsetto more closely and investigate meaning at the n + 1st order. Podesva (2013) outlines how although falsetto is not feature of African American English, Black speakers comment on it as a distinctly African-American way of speaking. He presents an example from one participant, who talks about how despite the oppression and the legacy of slavery, Black American culture is 'still there' in 'the way we practice religion, the way we talk to our friends, even- even in the way we talk', then uses falsetto in demonstrating this way of talking. He then considers the function of falsetto in more detail through the use of conversation analysis, and finds that falsetto tends to

be used to produce a negative stance - often, to negatively evaluate gentrification, by speakers who have strong community ties. Podesva (2013) argues that these speakers use falsetto as a linguistic act of resistance.

Similarly, Yuasa (2010) combines quantitative analysis of creak usage with a perception study to consider gender-related meanings of creaky voice in Japanese and American English. She considers first how rates of creaky voice used by young English-speaking American women compare to male counterparts and to female Japanese speakers, as well as what associations American listeners have with creaky voice as used by young female American English speakers. As well as finding that American women creaked more than their male and Japanese-speaking counterparts, she found that American listeners attributed a range of meanings to creaky voice in the speech of American women, including hesitant, non-aggressive, and informal, educated, urban-oriented, and upwardly mobile.

Similarly, Podesva (2018) looks at not only who creaks, but what affect they express with it, in a study of 42 American English speakers in interaction with video data that allows quantification of smiling and bodily movement. He finds that speakers creak less when their bodies are more animated, and that female speakers creak less when they are smiling. He argues that creak is therefore exploited by all speakers to convey a negative, disengaged stance, rather than being directly related to speaker gender.

3.3 Iconic: Iconic links with pitch and iconisation of variation

Not all ways that the meaning of voice quality is interpreted by listeners can be fully explained in terms of indexical meaning, however. As D'Onofrio & Eckert (2021) explore, signs can be described as iconic when they are construed as resembling their object in some way. In this way, the low f0 of creak can be interpreted as resembling masculinity through a link to the lower pitched voices that adult male speakers tend to have in comparison to female speakers (Podesva & Callier 2015: 115).

Associations of this kind are sometimes framed as having a biological basis. Ohala (1994) presents his theory of the frequency code in an attempt to explain cross-linguistic tendencies in the meaning of high and/or rising and low and/or falling intonation patterns. He suggests there are cross-cultural similarities in the use of f0 to signal affect, with high f0 corresponding to deference, politeness, submission and lack of confidence, but low f0 corresponding to assertiveness, authority, aggression, confidence and threat. Ohala (1994) argues that these tendencies are present throughout animal vocalisation, linguistic tones, and the use of vowel and consonants in words that denote

or connote smallness and largeness. Further, he argues that sexual dimorphism in vocal anatomy demonstrates that this variation is innate: that enlargement of the larynx due to exposure to testosterone is a product of evolution that 'occurs to enhance the acoustic component of aggressive displays' and occurs when 'males' are 'ready to compete for and retain the favors of a female'. He argues that this in turn suggests that there is an innate predisposition for listeners to recognise the meaning of different frequency sounds.

On the other hand, Eckert (2017: 1199) argues that while links between f0 and size are common to many languages, the idea that this is natural or universal is 'overstatement at best'. Instead, she argues that relationships between size and frequency are conventionalized when they brought into language, giving speakers and languages the ability to use sound to engage affectively with smallness and largeness, with a range of positive and negative qualities that are not set in stone. These potential meanings are underspecified, she argues: Instead of there being a fixed relationship between an particular iconic form and its meaning, the larger stylistic and interactional context determines how listeners interpret a link.

Icons can further be seen as a result of a process of iconization, one of the semiotic strategies that Irvine & Gal (2000) identify as a way that language ideologies explain sociolinguistic variation, naturalising links between groups of speakers and certain linguistic forms. Through iconiziation, they argue that the contextual relationship between an index and social group is reinterpreted as an iconic one, where something about a linguistic form is seen as resembling a social group that is considered to use that linguistic form.

Jeong (2017) explores the construction of iconic meaning with regards to voice quality in an exploration how three classic Hollywood archetypes - the dumb blonde, the femme fatale and the screwball heroine - are constructed through pitch and voice quality. She finds that femme fatales are constructed with the use of low pitch and low f0 variation, while the dumb blonde archetype is characterized by high pitch and high f0 variation. In each of these cases, she argues, the use of pitch is iconic: For the femme fatales, low pitch is linked to the authority and power that they exert over male protagonists, while the low pitch variability is used to construct a calm, unperturbed character, and link that is fortified through the dark lighting used in film noir. Meanwhile, the use of high pitch by dumb blondes is iconically linked to childlike behaviour through a connection to small size and higher pitch used by children, and to naively enthusiastic characterisation through bright, warm lighting. Both archetypes are linked by their use of breathy voice, which in turn comes to index the highly sexualised nature of these character types. Through the stylistic bundling of these variables and non-linguistic resources, such as lighting and visual imagery, the use of these linguistic features to portray these archetypes naturalizes the link between them

through iconization. The use of the variables in film then allows the social meaning of these variables to continue circulating in wider culture and allows speakers to draw on them in the construction of female stereotypes. The iconization and naturalisation of a link between breathy voice and sexuality is evident when we return to the sociophonetic literature: Recall Henton & Bladon's (1985) explanation of women's use of breathy voice in terms of speculation about women imitating of a state of vocal fold lubrication that they hypothesize could occur during sexual arousal.

Iconic links between creaky voice and masculinity through low fundamental frequency are often drawn on when attempting to explain the use and evaluation of creaky voice. Yuasa (2010: 331-332) suggested that female speakers may be using creaky voice to lower their voices, enabling them to 'project the image of educated urban professional women capable of competing with their male counterparts'.

The Effort Code (Gussenhoven 2002) is also sometimes alluded to in order to explain speakers' use of voice quality. The Effort Code proposes that increases in effort used to produce speech result in greater pitch variation, which can in turn be interpreted by listeners to as conveying things like emphasis or surprise. Yuasa (2010) suggests that the fact that creaky voice is produced with non-forceful airflow leads to impressions of hesitancy and non-aggression by listeners. Podesva (2013) draws on the Effort Code to explain the use of falsetto by Black American women, arguing that falsetto is exploited as a linguistic form of power.

Stross (2013) attempts to reconcile more biologically-focused explanations for iconic links proposed by Ohala (1994) and Gussenhoven (2002) and more culturally-mediated explanations (Mendoza-Denton 2011, Podesva 2007). Focusing on the way that falsetto can take on a range of meanings, he argues that whether these codes are biological is a relatively unimportant question. Instead, he argues that there are certain inferences and assumptions that are shared by many humans - what he terms 'observational logic' - because of our common experiences as humans in the world with similar sensory receptors. Following observational logic, he argues, allows for the fact that many interpretations of falsetto and high pitch are similar across languages and cultures, but that specific meanings of falsetto not determined by this and are embedded in specific cultural contexts.

3.4 Summary

Variation in voice quality can, then, take on a wide range of meanings, some of which exist as macro-level relationships between voice quality and categories like gender, class and ethnicity, and some which serve interactional purposes such as stance-taking, expression of affect, or persona management. These different levels of meaning are

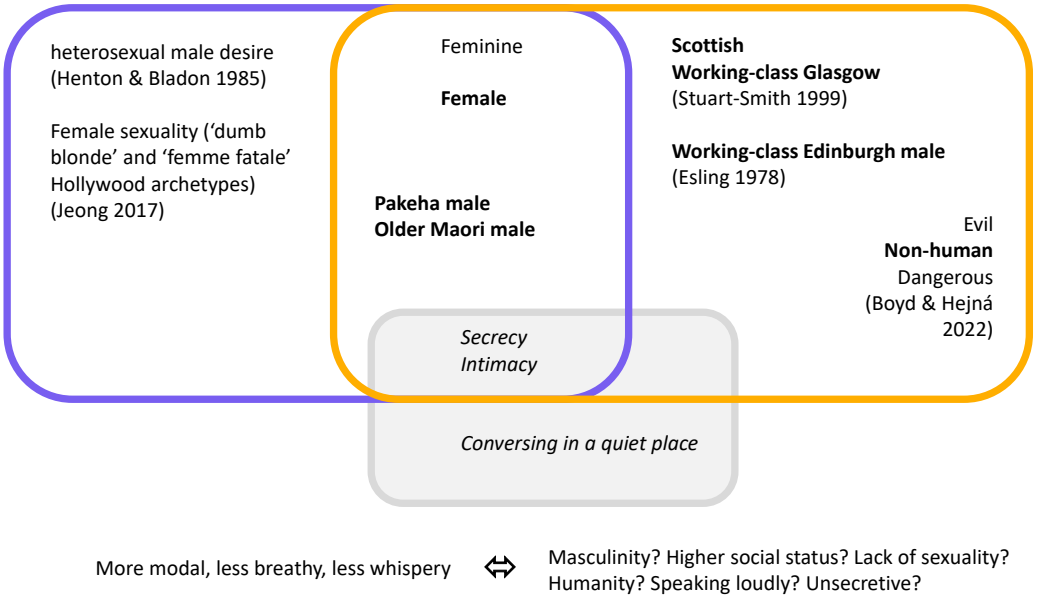


Figure 3.2: A representation of the overlapping indexical fields of different phonation types

connected by the orders of indexicality: A temporary stance can be re-interpreted by listeners, popular imagination, media and researchers as being connected to a macro-level social category. These meanings can be activated in interaction when they are used alongside other relevant variables - creaky voice may index masculinity in Modified Northern male speakers, but a very different picture emerges when it is used by a young female speaker in California alongside uptalk. A non-exhaustive representation of some of these possible social meanings of different voice qualities is presented in Figure 3.2, where temporary stances are represented in italics, qualities that are seen as being more permanent in normal text, and social types are identified in bold. These indexical meanings are also mediated in turn by iconic associations between particular voice qualities and social traits, and by the potential for relationships between voice qualities and speaker groups to be iconized as listeners see a speaker's voice quality as resembling some innate quality within them.

Chapter 4

Measuring voice quality

In Chapter 2, I introduced theoretical understandings of voice quality and gave an overview of descriptive categories for voice quality. Researchers take a variety of approaches to to studying voice quality, and each has different advantages and limitations, is appropriate to different research aims, and relates to the researcher's theoretical perspective. Methods of analysing voice quality can be separated into three main categories:

- Auditory-perceptual methods, which involve listening to a voice and following a protocol to make systematic decisions about its quality (e.g. Laver et al. 1991[1981])
- Acoustic methods, which involve measuring the properties of the speech signal such as periodicity and spectral tilt to infer properties of the auditory quality of the voice (e.g. Kreiman, Gerratt & Antoñanzas-Barroso 2007)
- Articulatory methods, which involve using an instrument such as a laryngoscope or an electroglottograph to investigate the physiological mechanism that underlies the auditory quality and acoustic signal of a voice (e.g. Esling 2005)

In the present study, I combine auditory-perceptual and acoustic methods, and therefore leave articulatory methods aside in the present discussion. In Part II, I conduct a corpus investigation into variation in voice quality in Scotland according to macro-level social categories as well as linguistic factors. In this Chapter, I give an overview of the different types of auditory-perceptual and acoustic methods that can be used to describe and quantify phonation. I present this overview with a view to developing an integrated auditory-perceptual and acoustic approach to analysing phonation in the present study. I aim to select methods that are appropriate to a corpus analysis of voice quality and which allow consideration of how short-term fluctuations in phonation within a speaker compound to produce overall laryngeal voice quality.

4.1 Auditory-perceptual methods

Auditory-perceptual approaches to analysing the voice involve listening to samples of a voice and following a descriptive system or protocol of some kind with the aim of systematically describing its auditory quality. This allows the auditory quality of a sample to be communicated to others in writing, and compared either to other samples produced by the same speaker other speakers, which can give insight into the function or meaning of a certain voice quality.

Certain auditory-perceptual methods are better suited to clinical, phonological or sociolinguistic applications. In clinical settings, a clinician's training background and the purposes of the assessment may affect whether the clinician uses a scale that focuses exclusively on phonation, such as the Grade, Roughness, Breathiness, Asthenia and Strain scale (GRBAS) (Hirano 1981), or a scheme that considers voice quality beyond phonation, like the Buffalo III Voice Profile (Wilson 1987) whose scope also includes breath supply and nasal resonance. Meanwhile, a researcher interested in phonation contrasts in a language may consider the vowels or consonants relevant to that contrast and conduct an auditory assessment of where a phonation category falls according to the glottal stricture model (Ladefoged 1971, Ladefoged & Maddieson 1996). Many sociolinguistic studies are interested in only certain phonation types, such as creak and falsetto (Podesva 2007). Examinations of this kind may be interested in the social meanings attached to voice quality, and often require analysis at the level of the vowel, syllable, or word, to examine how phonation varies according to linguistic or contextual factors. Across each of these domains, Voice Quality Symbols (Ball, Esling & Dickson 1995) may be used to transcribe voice quality.

Vocal Profile Analysis (VPA) (Laver et al. 1991[1981]), which was developed alongside Laver's (1980) systematic description of voice quality, has been used in both clinical (e.g. Beck 1988) and sociolinguistic (e.g. Esling 1978b) investigation of voice quality, as well as for forensic purposes (Segundo & Mompean 2017). VPA is a protocol for analysing a speaker's 'vocal profile', defined by Laver et al. (1991[1981]: 265) as 'a statement of the speaker-characterising, long-term features of a person's overall vocal performance'. In contrast to scales like GRBAS that only consider phonation, VPA is a componential approach involving a comprehensive evaluation of the overall perceptual quality of a speaker's voice. This includes both laryngeal and supralaryngeal settings, as well as other suprasegmental features such as pitch, loudness and speech rate. Different settings or components, like 'breathiness', can be evaluated in terms of their neutrality or non-neutrality, and where non-neutral, can be rated on scalar degrees. Together, the 'constellation' of settings used by a speaker makes up their vocal profile, a phonetic summary of a speaker's habitual 'voice' (Laver et al. 1991[1981]: 265). In his initial presentation of what would later be known as the Vocal Profile Analysis,

Vocal Profile Analysis Protocol

Judge: ..... Tape: ..... Sex: .....
Date of Analysis: ..... Speaker: ..... Age: .....

I. VOCAL QUALITY FEATURES

CATEGORY	FIRST PASS: Neutral	FIRST PASS: Non-neutral: Normal	FIRST PASS: Non-neutral: Abnormal	SECOND PASS: SETTING	SECOND PASS: Scalar Degrees: Normal 1	2	3	Abnormal 4	5	6
A. Supralaryngeal Features										
1. Labial				Lip Rounding/Protrusion						
				Lip Spreading						
				Labiodentalisation						
				Extensive Range						
				Minimised Range						
2. Mandibular				Close Jaw						
				Open Jaw						
				Protruded Jaw						
				Extensive Range						
				Minimised Range						
3. Lingual Tip/Blade				Advanced						
				Retracted						
4. Lingual Body				Fronted Body						
				Backed Body						
				Raised Body						
				Lowered Body						
				Extensive Range						
				Minimised Range						
5. Velopharyngeal				Nasal						
				Audible Nasal Escape						
				Denasal						
6. Pharyngeal				Pharyngeal Constriction						
7. Supralaryngeal Tension				Tense						
				Lax						
B. Laryngeal Features										
8. Laryngeal Tension				Tense						
				Lax						
9. Larynx Position				Raised						
				Lowered						
10. Phonation Type				Harshness						
				Whisper(y)						
				Breathiness						
				Creak(y)						
				Falsetto						
				Modal Voice						

(a) Section I of the protocol

II. PROSODIC FEATURES

CATEGORY	FIRST PASS: Neutral	FIRST PASS: Non-neutral: Normal	FIRST PASS: Non-neutral: Abnormal	SECOND PASS: SETTING	SECOND PASS: Scalar Degrees: Normal 1	2	3	Abnormal 4	5	6
1. Pitch				High Mean						
				Low Mean						
				Wide Range						
				Narrow Range						
				High Variability						
				Low Variability						
2. Consistency				Tremor						
3. Loudness				High Mean						
				Low Mean						
				Wide Range						
				Narrow Range						
				High Variability						
				Low Variability						

III. TEMPORAL ORGANISATION FEATURES

CATEGORY	FIRST PASS: Adequate	FIRST PASS: Inadequate	SECOND PASS: Scalar Degrees of Inadequate: 1	2	3	
1. Continuity						Interrupted
2. Rate						Fast
						Slow

IV. COMMENTS

	Adequate	Inadequate	1	2	3
Breath Support					
Rhythmicality					

	Present	Absent
Diplophonia		

'Vocal Profiles of Speech Disorders' Research Project. (MRC Grant No. G978/1192) Phonetics Laboratory, Department of Linguistics, University of Edinburgh. © June 1981.

(b) Section II, II, IV of protocol

Figure 4.1: The Vocal Profile Analysis Protocol from Figure 15.1 from Laver et al. (1991[1981]: 268)

Laver (1980: 157-162) summarised the settings discussed in previous chapters of the book, listing which settings could be compatible, and introducing the idea that these settings could be rated on scalar degrees.

Figure 4.1a and Figure 4.1b show a blank Vocal Profile Analysis protocol form which can be used in the analysis of a speaker's vocal profile, as presented in Laver et al. (1991[1981]).

One distinctive feature of VPA is the existence of a 'neutral' setting which exists a standard reference setting with a clearly defined auditory quality and underlying articulatory configuration that allows voice quality to be described in relation to a clear reference point (Beck 2005: 296). Laver (1980: 14-15) emphasises that the neutral setting is not a description of what is considered to be 'normal' or a description of the rest position of the vocal tract. For phonation, this neutral setting is modal voice; following van den Berg (1968), Laver (1980: 14) defines modal voice as the state where 'the vibration of the true vocal folds is regularly periodic, efficient in air use, without audible friction, with the folds in full glottal vibration under moderate longitudinal tension, moderate adductive tension and moderate medial compression'.

In the process of completing a VPA protocol, a trained judge will first complete a 'first pass', marking whether the voice is neutral or non-neutral in each category. Following this, they complete a 'second pass', where they judge deviation from the

neutral setting in terms of scalar degrees (Laver et al. 1991[1981]: 270-271). Scalar degrees 1-3 are considered to be 'normal', while scalar degrees 4-6 are considered to be 'abnormal', a distinction which Laver et al. (1991[1981]: 271) recognise as 'somewhat problematic' due to a lack of information about the distribution of settings in non-pathological voices. In practical terms, a setting rated as 'abnormal' is considered to be one which requires treatment, though they recognise that the way settings relate to each other and a patient's own assessment of vocal function is relevant to this as well (Laver et al. 1991[1981]: 271-272).

In this research, I draw extensively from VPA in the development of my own auditory-perceptual analysis scheme, termed Phonation Profile Analysis (PPA). I take VPA as the starting point because it is grounded in the phonetic description of voice quality (Laver 1980) and has previously been used in previous analyses of sociolinguistic variation in Scottish voice quality (Esling 1978b, Stuart-Smith 1999b, Beck & Schaeffler 2015). However, the research presented here diverges from the aims of VPA, necessitating the development of a novel approach adapted to these purposes. Firstly, the present research focuses solely on laryngeal voice quality. Furthermore, I aim to consider how short-term fluctuations in quality compound to produce longer-term perceptual quality, specifically considering variation according to linguistic factors. Both of these changes require VPA to be adapted in a number of ways.

In the rest of this section, I begin by discussing what auditory-perceptual analysis can bring to the present study. I then describe previous adaptations of VPA, detail the principles of VPA maintained in PPA, and then describe points of divergence. Finally, I consider the limitations of auditory-perceptual analysis. In Section 4.2, I turn my focus to a background of acoustic methods.

4.1.1 Why conduct auditory-perceptual analysis?

4.1.1.1 Ease of interpretation

Laver (1974) outlined a large number of impressionistic labels for voice quality that exist in lay terms, including both labels describing auditory vocal quality (e.g. 'raspy', 'gravelly') and those that make reference to a speaker's imputed character (e.g. 'a man's voice', 'an authoritative voice'). These labels are ambiguous in terms of their phonetic quality, but reveal the great amount of indexical meaning that voices can carry. If we wish to study these indexical meanings, it makes some sense to begin in the auditory-perceptual domain, where any variation in quality that relates to indexical meaning should be perceptible in some way. Systematic auditory-perceptual analysis of voice quality allows us to capture the variation in voice quality and describe it in terms that can be understood by others.

In comparison to acoustic methods, auditory-perceptual methods therefore have the advantage of being clearer to interpret. While the auditory effect of a difference in acoustic measurement may not always be apparent, auditory-perceptual methods are regarded as the 'gold standard' that acoustic measurements are compared against (Segundo & Mompean 2017). As discussed further in Section 4.2.6.1, there is no straightforward way to interpret a one-unit difference in an acoustic measurement as a difference in voice quality, and instead acoustic measures need to be interpreted relative to each other and in tandem with other measures.

Certain phonation types are also potentially more easy to identify auditorily than using purely acoustic methods. As Boyd & Hejná (2022) discuss, breathy voice and whispery voice can be differentiated auditorily, and can serve different purposes, but are often difficult to differentiate acoustically.

4.1.1.2 Possibility to combine with qualitative acoustic analysis

Auditory analysis can also be combined with supporting analysis of the spectrogram or spectral slices. Wileman (2018: 110) notes that creaky voice is often considered relatively easy to identify auditorily with help from spectral cues. Figure 4.2 shows an example of creaky voice from Yuasa (2010), where the presence of creak was coded at the level of the word using auditory criteria and supporting evidence from the spectrogram and waveform. As creak is often visible with dark vertical striations in the spectrogram and dampened, far apart, irregular glottal pulses in the waveform (Figure 4.2), the presence of creak in a vowel, syllable or word can be straightforward to identify.

Moisik (2013) also demonstrates the value of combining auditory judgements with qualitative acoustic analysis. He considers harsh voice in popular American media, and currently no automated method for identifying harsh voice exists, necessitating an auditory approach. Auditory identification of harsh voice allows him to consider how harsh voice is used to convey aggression and heightened emotional states and its role in the portrayal of racial stereotypes of black people. Finally, he considers the acoustic form of instances of harsh voice that are identified auditorily, giving insight into its spectral qualities which included subharmonic content, interharmonic noise, and increased acoustic energy above 1,000 Hz.

4.1.1.3 Breadth and depth of analysis

Auditory-perceptual methods also have the potential to speak to the multi-dimensional nature of voice quality. As Beck (2005: 291) notes, with a method like VPA, it is possible to do a broad analysis of a speaker's whole production system, separated out

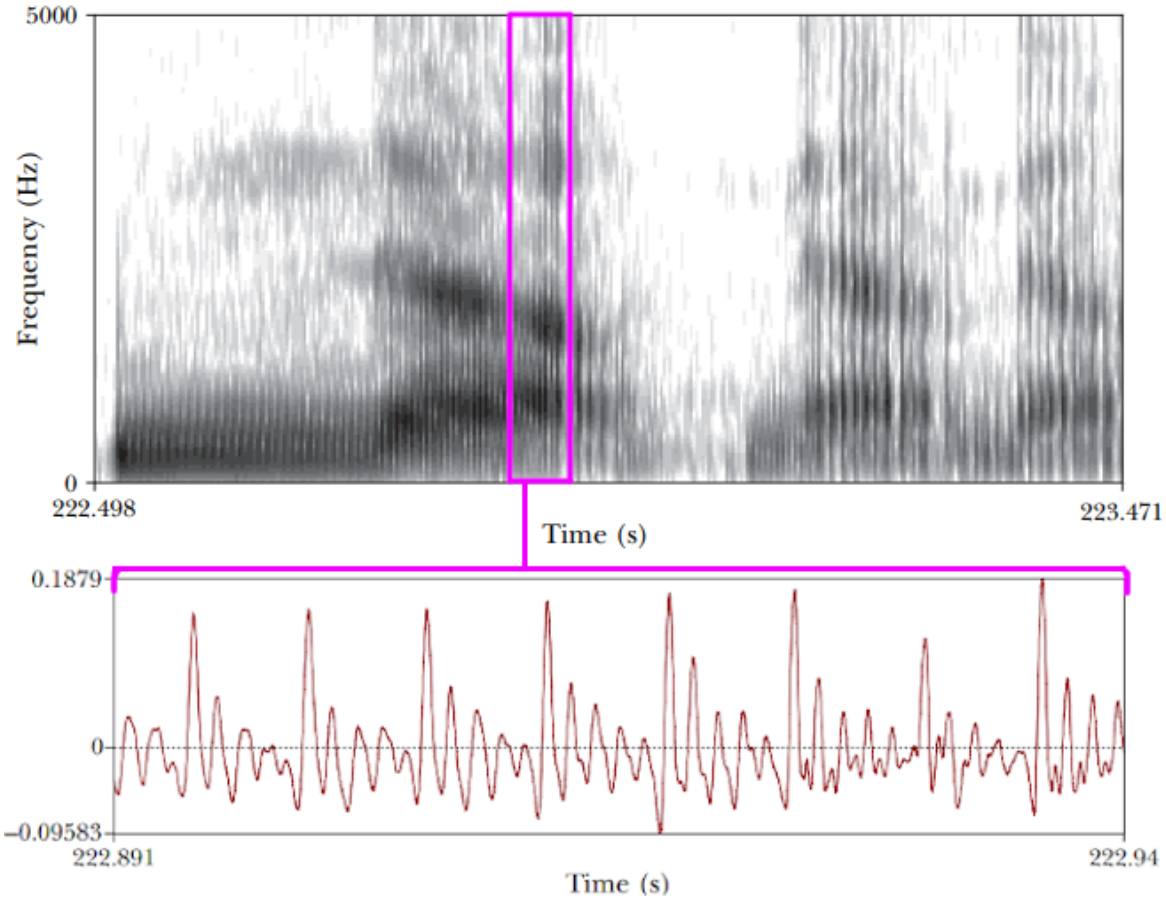


Figure 4.2: Spectrogram and waveform for the utterance "No, no, no" by an American female informant taken from Fig. 1 and Fig. 2 in Yuasa (2010: 324-325), showing vertical striations and irregular glottal pulses that aid coder in identifying precise location of creak onset. My own annotations in pink show the approximate part of the spectrogram that has been expanded into the waveform.

into the different components that constitute the resulting auditory quality. Though time-consuming, the act of spending a large amount of time listening to the voice of a single speaker or small number of speakers in this way allows for deeper engagement with a speaker's voice qualities in a way that would be difficult in an automated acoustic approach, resulting in increased breadth and depth of analysis.

4.1.2 From Vocal Profile Analysis to Phonation Profile Analysis

4.1.2.1 Previous adaptations and alterations of VPA

My adaptation of VPA involves restricting its focus to phonation and working with short stretches to allow considerations of short-term variation in phonation. However, this is by no means the first attempt to adapt VPA for specific purposes. As Beck (2005: 295) emphasises, VPA is best understood as an approach to the analysis of

voice quality, rather than as a single procedure; the protocol has evolved to reflect changes in understanding of voice quality, adapted in light of what has been learned from the experience of applying it, and adapted to suit different purposes.

Underlying each iteration of the VPA protocol are a number of general principles that differentiate it from other approaches. Beck (2005) summarises these principles:

1. Voice quality considers the whole of the vocal apparatus as contributing to an individual's characteristic voice quality
2. Voice quality is treated as being composed of different settings that can be combined in differed ways
3. Voice quality is not compared to a 'normal' baseline, but to a 'neutral' setting
4. Deviations from the neutral setting are quantified on scalar degrees
5. Differences in voice quality resulting from anatomy and phonetic adjustments made by the speaker may sound auditorily equivalent, so are not distinguished

In some instances, changes resulted from experience applying the protocol. Figure 4.1a and Figure 4.1b presents the tenth version of the form, which took shape through collaboration between the authors of Laver et al. (1991[1981]) and experienced speech therapists over the course of VPA training sessions. This iteration contains six degrees for each scalar setting. Beck (2005: 299) notes that this was later reduced for several settings: She notes that new judges often made errors with such a fine-grained scale of variation, while experienced judges found that the scale of variation was best captured by fewer degrees.

The version presented in 4.3 shows a later iteration of the protocol with fewer scalar degrees from Laver (1994: 154). In addition to this, changes to Laver's conceptualisation of voice quality are reflected here, with breathy and harsh voice moved to the 'Laryngeal Tension' section.

While initially the protocol was developed with a focus on clinical usage, it has a wide variety of applications. One application is in sociolinguistics, to examine variation in voice quality settings between different populations of speakers. This was the case in Beck & Schaeffler (2015)'s study of voice quality in adolescents in three locations in Scotland. They adapted VPA to reflect the sociolinguistic focus of the study: They maintain the 6-degree scheme shown similar to that shown in Figure 4.1, but term scalar degrees 1-3 'moderate' rather 'normal' and 4-6 as 'extreme' rather than 'abnormal', reflecting the fact that the protocol is not being used in a clinical setting. Furthermore, Beck & Schaeffler (2015) and Stuart-Smith (1999a) only record voice quality features using Section 1 of the protocol, and do not investigate prosodic

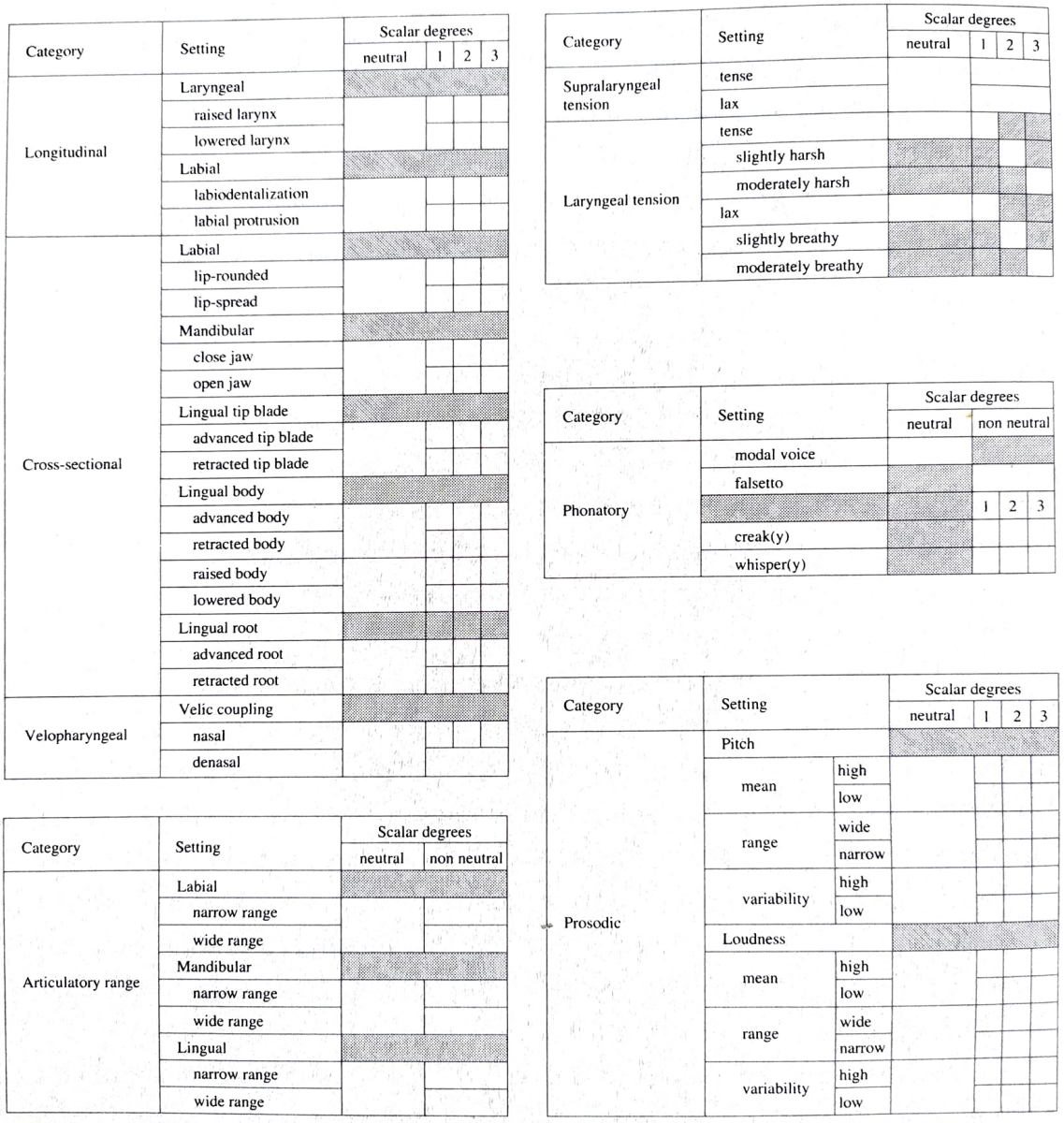


Figure 4.3: A later version of the protocol presented in Figure 5.14 in Laver (1994: 154)

or temporal features which are outside the scope of their research. Schaeffler, Eichner & Beck (2019) restrict their focus further to harsh, creakiness and whisperiness in a study on whether acoustic measurements can be used to classify voices according to VPA labels. These are important shifts away from VPA's original focus on describing a person's entire vocal profile, towards only those relevant for a particular study.

Furthermore, Segundo & Mompean (2017) developed a simplified VPA (SVPA) protocol with the aim of facilitating assessment of voice quality and speaker similarity for forensic application. They noted that raters often find it hard to isolate related settings and sometimes disagree on the definitions of labels, and suggested that a simplified protocol might suffice for non-clinical settings. Segundo & Mompean (2017) simplified the original VPA protocol by reducing the number of settings, proposing a protocol with 10 groups of settings and two possible articulatory strategies for deviating

from neutral for each. For example, rather than rating each possible phonation type on a scale of 1-6, raters instead rated laryngeal irregularity, which if non-neutral, could be either harsh or creaky, and laryngeal friction, which could be either breathy or whispery. They investigated how reliable the SVPA was in terms of inter- and intra-rater reliability and found that while reliability between two phonetically trained raters varied by setting, raters were found to be highly consistent internally.

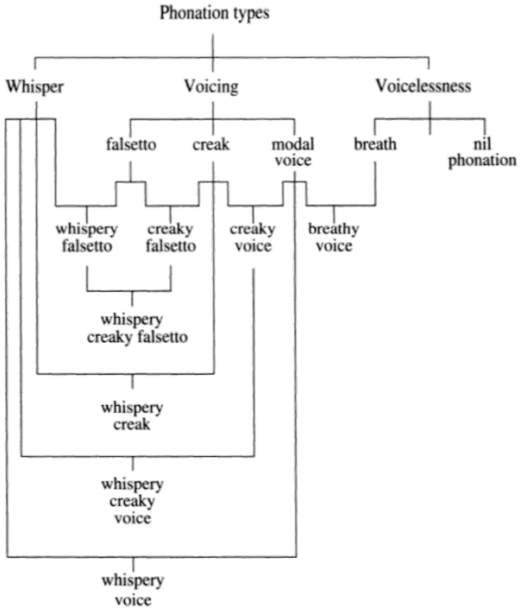


Figure 4.4: Figure 7.7 from Laver (1994: 199) Constraints on the combinability of different modes of phonation

4.1.3 From VPA to PPA: Points of similarity

4.1.3.1 Maintaining VPA terminology and categories

Different studies operationalise different distinctions between different phonation types. Laver et al. (1991[1981]: 213) notes that textbooks use terms like hoarse, husky and rough to describe the quality created by certain vocal fold problems, but that it is impossible to tell whether these terms are equivalent, creating a situation where 'a single label has multiple potential referents, and a single phenomenon has multiple potential labels'.

For example, Podesva (2013) makes the distinction between breathy voice and whispery voice on the basis of voicing, coding syllables all syllables with aspiration noise and voicing as breathy, and syllables that do not occur with voicing as whispery voice. This approach may reflect a potential difficulty of differentiating breathy and whispery voice when listening at the level of the syllable. However, as the distinction between breathy voice and whispery voice differs from that of studies that use VPA, and it is

difficult to tell whether the 'breathy voice' in Podesva (2013)'s study is equivalent to that in Stuart-Smith (1999b), for example, making comparison across studies difficult.

This is particularly apparent for creak, with Keating, Garellek & Kreiman (2015) differentiating prototypical creak, vocal fry, multiply pulsed creak, aperiodic creak, and tense voice on the basis of acoustic and auditory qualities, but other researchers considering grouping some or all of these qualities as part of the same phenomenon.

In PPA, I aim to stay as close as possible to VPA in defining phonation types. I therefore maintain modal voice as a neutral baseline (discussed further in Section 4.1.3.3) and distinguish between breathy and whispery voice. This maintains a degree of comparability between the present study and previous VPA research on Scottish accents (Esling 1978b, Stuart-Smith 1999b, Beck & Schaeffler 2015). However, when defining terms, I also refer to Esling et al. (2019), who draws on Laver (1980), to ensure that PPA incorporates developments from the Laryngeal Articulator Model.

More information about how these categories are defined in the present study is given in Section 6.2, but to give an overview, PPA includes modal voice, whisper, falsetto, harsh voice, creaky voice, breathy voice and whispery voice. This includes lax voice as the lowest scalar degree for breathy voice, and tense voice as the lowest scalar degree for creaky voice.

4.1.3.2 Voice quality as componential

In PPA, I maintain VPA's approach to analysing voice quality as componential and do not restrict my analyis to a single phonation type. In Laver's (1980, 1994) understanding of voice quality, the perceptual effect of phonation is the product of different settings of the vocal tract, some of which can occur alone as simple phonation types (modal voice, falsetto, whispery, creak), and others which are the product of different settings combining (e.g. harsh whispery creaky falsetto). PPA will follow VPA in taking this as its theoretical underpinning, and consider that the perceptual quality of a voice may be result of combinations of different settings.

This comes in contrast to many sociolinguistic analyses of phonation, which often restrict their scope to a smaller number of phonation types and do not consider combination types. This can be useful in cases where the sociolinguistic function of a single setting is analysed in greater detail, as has been the case with creaky voice (e.g. Yuasa 2010, Becker, Khan & Zimman 2022). However, previous VPA descriptions of Scottish voice quality reveal that combinations between different settings, such as tense whispery (Stuart-Smith 1999b) or harsh whispery voice (Esling 1978b) pattern according to social grouping, so including combination types appears essential to the present research.

The process by which combinatorial possibilities were decided is detailed in Section 6.2, but here I give a brief overview. I treated modal voice and whisper as unable to combine with other categories. I treated all states that involve laryngeal constriction (Esling et al. 2019), whispery voice, creaky voice and harsh voice as able to combine with each other. Contrary to Laver (1980), but in line with more recent research that reveals that creaky voice can be produced with unconstricted configuration (Gobl & Chasaide 2000, Slifka 2006, Keating, Garellek & Kreiman 2015), I treated breathy voice and creaky voice as able to combine.

4.1.3.3 Phonation is compared to modal voice

In PPA, I maintain Laver: 14-15, 94's (1980) narrow definition of modal voice in terms of auditory, acoustic and articulatory criteria.

In sociolinguistic studies that do not use VPA, there are many other definitions of modal voice. In studies that only consider creaky voice, modal voice often becomes anything that is not creaky (e.g. Yuasa 2010). Meanwhile, Podesva (2013) defines it only as 'unmarked' voicing, and his findings that modal voice is more common than any other phonation type suggest that he may have operationalised modal voice as being a speaker-specific baseline, rather than following the strict definition that underlies VPA. Wileman (2018: 112) draws on Gerratt & Kreiman (2001: 377) when discussing how differentiating between modal and breathy voice is difficult because of the fact that these form a continnum with no clear cut-off. Because of this, Wileman (2018: 113) decides to include segments that display varying degrees of breathiness within the category of 'modal'.

In the present study, I attempt to ensure consistency with previous VPA research by treating modal voice as a neutral baseline category, to which other phonation types are compared. Following Laver (1980: 94) modal voice is taken as a state where 'the vibration of the true vocal folds is periodic, efficient and without audible friction'.

4.1.3.4 Voice quality can be rated on scalar degrees

The ability for deviations from the neutral baseline to be quantified on scalar degrees is a key feature of VPA, with different iterations using either 3 or 6 degrees. Outside of VPA research, sociolinguistic auditory-perceptual methods rarely use scalar degrees (See, for example, Yuasa (2010), Wileman (2018), Podesva (2007, 2013), Becker, Khan & Zimman (2022)).

In PPA, I follow VPA in allowing whispery voice, breathy voice, and creaky voice to be rated on scalar degrees. However, as discussed below in Section 4.1.4.2, PPA

diverges from VPA in the number of scalar degrees used.

4.1.4 From VPA to PPA: Points of divergence

As previously noted, PPA's focus is restricted to laryngeal voice quality; supra-laryngeal and prosodic aspects are not considered. This allows more detailed analysis of laryngeal components. Here, I outline how PPA diverges from VPA and illustrate these differences with reference to a worked example.

4.1.4.1 Overall laryngeal voice quality is made up of shorter-term variation in phonation

The restriction to laryngeal components allows consideration of variation at a shorter time scale. Consider a hypothetical situation where you are analysing a one-minute extract of two speakers with whispery voices. Speaker A uses a moderately whispery voice consistently throughout the conversation across each utterance evenly, without much variation away from this mode. Speaker B fluctuates between slightly whispery, moderately whispery, and extremely whispery at different points throughout a one minute sample. These differences would be hard to capture in VPA, and analysis of both speakers would likely show moderate non-pathological use of whispery voice in both speakers. PPA however aims to be able to track where the second speaker becomes more or less whispery. Variation in voice quality could then be linked to conversational functions: Telling a secret, or voicing a character in constructed dialogue, for example.

If we wish to understand a speaker's longer term laryngeal voice quality in terms of shorter term fluctuations, we need to choose a new unit of analysis. Drawing on Laver's concept of susceptibility (Laver 1980: 20-21), I argue that vowels and sonorant consonants are the most susceptible to changes in phonation. I therefore take stretches of sonorants as the unit of analysis.

A worked example of this is shown in Figure 4.5. This utterance consists of 7 stretches of sonorants, which I terms 'voiced stretches', shown on Tier 1.

4.1.4.2 A five-point scale for scalar degrees

VPA typically involves rating settings on 6-point scales, where degrees 1-3 are 'normal' or 'moderate' and degrees 4-6 are 'abnormal' or 'extreme'. This distinction, though not unproblematic, is clinically useful as a way of distinguishing non-pathological use (e.g. an accent feature) from more extreme use symptomatic of a voice disorder.

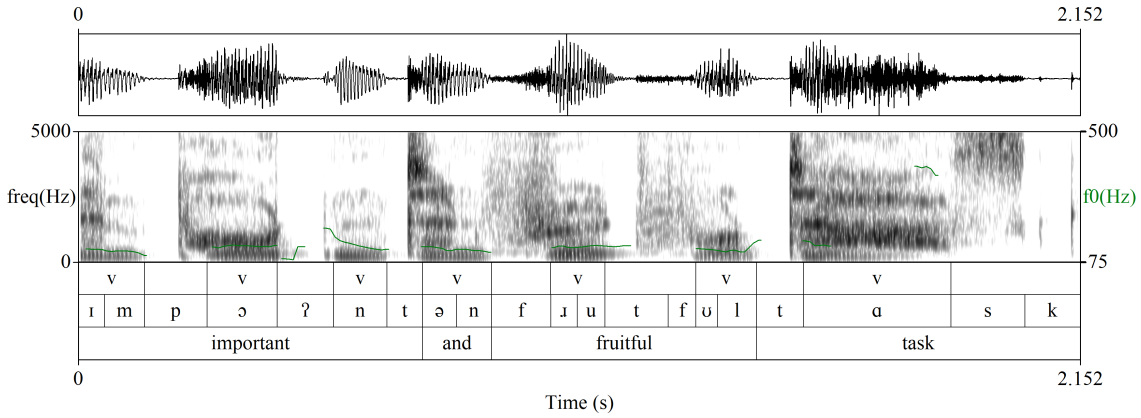


Figure 4.5: John Laver producing 'important and fruitful task'. Voiced stretches coded as 'v' on Tier 1.

However, PPA's sociolinguistic focus means that scale with a perceptual midpoint would be more useful. As described later in Section 6.2.5.1, a number of different ordinal scales were tested, with the final version of PPA presented in this thesis using a 5-point scale. This allowed scalar degree 3 to function as a midpoint, diverging from VPA where the boundary from degree 3 to 4 represents a shift from normal to abnormal voice quality.

In line with VPA, modal voice, whisper and falsetto are treated as binary rather than rated on scalar degrees, while whispery voice, breathy voice and creaky voice are all rated on scalar degrees. However, PPA departs from VPA in treating harsh voice as binary rather than scalar, due to how rarely it appeared in my data.

Figure 4.6 shows an example of the same utterance shown in Figure 4.5, coded in PPA. While the whole utterance is whispery in some respect, PPA allows variation in the auditory quality between different stretches to be captured. For example, the [n] in 'important' contains more low-frequency, more-periodic energy in the spectrogram, and an auditorily less whispery quality, leading it to be coded only scalar degree 2. Meanwhile, the [ʊl] in 'fruitful' actually contains a degree of harshness, so this is also coded as harsh voice. The end of the utterance is so whispery that voicing disappears, leading the final stretch to be divided into two stretches to show the change in quality: The first part is coded as scalar degree 5 for whispery voice, while the final stretch is coded as scalar degree 1 for whisper.

This shows the value of PPA for representing shifts in voice quality. VPA would have necessitated choice of a single scalar degree, missing the short-term variation. Further, if we had coded each individual word or syllable solely for the presence or absence of whispery voice, we would miss the fact that 'task' is far more whispery than 'important', and that towards the end of the [a] of 'task', voicing stops and phonation becomes whisper rather than whispery voice.

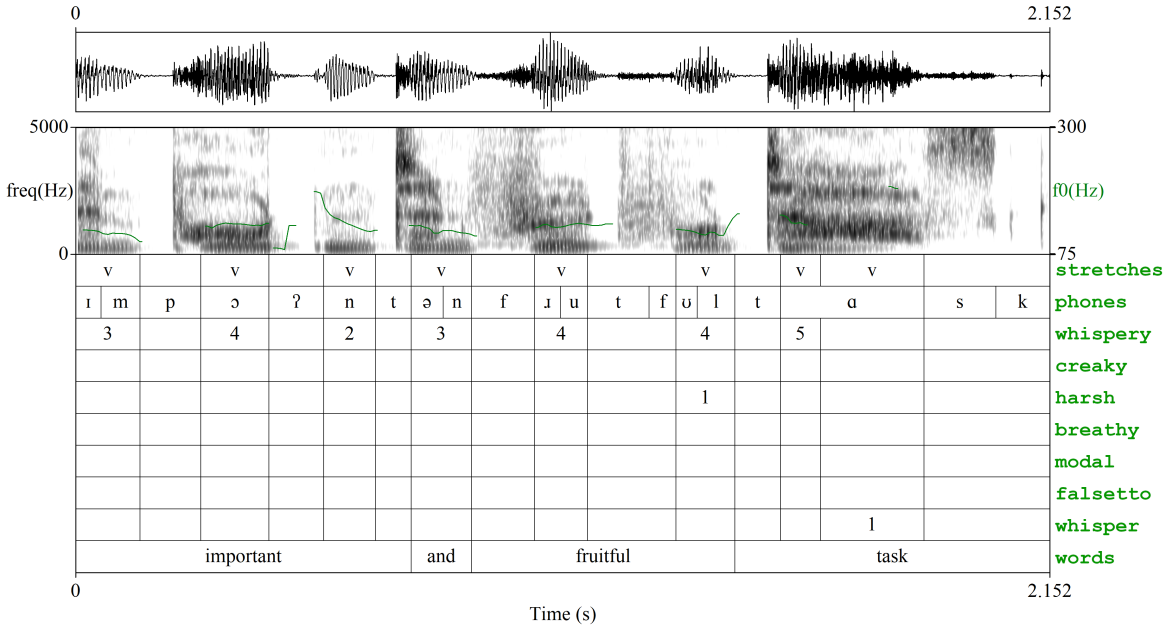


Figure 4.6: John Laver producing 'important and fruitful task' coded in PPA.

However, auditory-perceptual analysis is not without limitations; because of this, in the present study I combine it with acoustic analysis. In the next section, I turn to a background of acoustic methods for analysing voice quality.

4.2 Acoustic methods

4.2.1 Why conduct an acoustic analysis?

In Section 4.1, I discussed some of the benefits of conducting an auditory perceptual-analysis and introduced Phonation Profile Analysis, the auditory-perceptual method I use in this study. However, this approach is time-consuming to conduct, meaning that only 90 seconds of speech from 24 speakers from SCOSYA could be examined using this approach. To investigate variation on a larger scale, this research therefore takes a two-phase approach to acoustic analysis: Conducting acoustic analysis on the same data as PPA, then rolling out this acoustic analysis to more data from a larger number of speakers. In this section, I explore the limitations of auditory-perceptual analysis and how acoustic analysis can give more information about phonetic detail, increase reliability, and allow analyses to be conducted on a larger scale.

4.2.1.1 Intra-/inter- rater reliability

Auditory-perceptual analysis is often viewed as subjective, and has been criticised on the grounds of low inter-rater reliability. VPA training is required for listeners to be

able to use VPA effectively, but even following training, reliability between listeners varies by setting. Laver et al. (1991[1981]) finds that judges who were trained on a two-and-half-day programme performed poorly when rating larynx position and tongue position, averaging errors of more than two scalar degrees when rating voices for these settings, but were more successful in their ratings of other settings. Similarly, Webb et al. (2004) find that inter-rater reliability for judges trained on a four-day VPA course ranged from poor to good for different settings, with ratings for laryngeal settings typically being more reliable than for supra-laryngeal settings. Increased reliability for laryngeal settings is promising for the present research, but reliability of auditory-perceptual methods remains a point of caution.

Though as Beck (2005: 290) notes, operation of equipment and interpretation of the data still varies between researchers in acoustic analysis, acoustic analysis should be more consistent, reliable, and reproducible than auditory-perceptual analysis. In the present research, it was not possible to have the data coded by a second rater in order to check inter-rater reliability of PPA; acoustic analysis therefore supports the validity of the auditory-perceptual part of this research.

4.2.1.2 Holistic perception

Considering the reasons for low rater reliability in auditory-perceptual approaches can also shed light on what else can be gained from an acoustic analysis. Kreiman, Gerratt & Antoñanzas-Barroso (2007) investigated the reasons for poor inter-rater reliability in auditory-perceptual voice quality assessment in a series of experiments where listeners assessed breathiness in relation to different comparison stimuli. In one condition, listeners assessed breathiness through comparison to reference stimuli, matched to the voice being rated, which demonstrated different degrees of breathiness. In another condition, listeners assessed breathiness with reference to a generic set of comparison stimuli that were not matched with the voice being rated. In another condition, listeners were given no comparison stimuli and used only their internal standards. Kreiman, Gerratt & Antoñanzas-Barroso (2007) found that the probability of listeners agreeing in their assessment of voice quality increased when they were presented with voice-matched comparison stimuli, but decreased when they were only given generic comparison stimuli, when compared with the listeners who used only their internal standards. Kreiman, Gerratt & Antoñanzas-Barroso (2007) suggest that this means that listeners hear voices as a holistic pattern rather than breaking them down into component features, and have difficulty paying attention to only one dimension unless they have access to context of the wider pattern of a voice. This gives acoustic analysis an advantage over purely auditory-perceptual approaches: If listeners hear voices as an overall pattern and have difficulty breaking it down into its constituent components, acoustic analysis may be able to give more information about which aspects of

the signal contribute to the overall impression of a voice. For example, we know that creaky voices can be produced with different glottal configurations which have different acoustic features (Keating, Garellek & Kreiman 2015), and while listeners may be able to hear that a voice sounds creaky, acoustic analysis may allow greater insight into the fine-grained phonetic detail that distinguishes different types of creaky voice.

4.2.1.3 Continuous perception

Furthermore, Kreiman, Gerratt & Antoñanzas-Barroso (2007) asked listeners to rate breathiness with either ordinal 6-point scales or continuous scales. They found that that listener agreement increased when they rated voice quality on continuous scales rather than ordinal ones. In most auditory-perceptual rating systems of voice quality, ordinal scales are used for practical reasons, but the choice of the number of points on an ordinal scale is largely arbitrary and constrains the ability of listeners to rate voice quality accurately. As this finding demonstrates, detail about voice quality that exists in the phonetic signal is continuous, and this fine-grained level of phonetic detail that affects our perception of overall voice quality is difficult to quantify with purely auditory-perceptual analysis. Acoustic analysis can therefore allow us to quantify perception in a more detailed, continuous manner.

4.2.1.4 Time variation

Another limitation of auditory-perceptual analysis is that voice quality varies according to linguistic, paralinguistic and extralinguistic factors (Abercrombie 1967) and is therefore not constant over time. Some of these changes may be very brief, however, and an auditory-perceptual analysis needs to be conducted over units that are long enough for the listener to be able to quantify and rate voice quality. VPA, for example, requires samples of at least 40 seconds of speech (Beck 2005). Although settings can be marked as 'intermittent' in VPA, VPA cannot capture the frequency of intermittent use of a setting, nor its auditory degree. PPA allows greater consideration of time variation than VPA, but each voiced stretch still needs to be long enough that voice quality can be perceived in the stretch and 100 ms is taken as a cut-off point for this and so it does not give any insight into fluctuations in voice quality that are shorter than this. This 100 ms threshold was determined through trial-and-error during the development of the method and is largely arbitrary as there is no agreed threshold where a momentary fluctuation in phonation begins to contribute to perceived quality, and little is known about the relationship between stimuli length and reliability of listener judgments of phonation type.

4.2.1.5 Scale of analysis

Auditory-perceptual analysis can be very time consuming to conduct, especially on spontaneous speech and in large sets of data. This compounds with issues of inter-rater reliability, as lengthy analysis time decreases the amount of data that can be coded by multiple raters. Acoustic measurement, meanwhile, can be run on larger sets of data more quickly, increasing the breadth of the analysis.

4.2.1.6 Acoustic analysis in the present study

Overall, then, acoustic analysis can increase the scale and depth of analysis, shedding light on how multiple acoustic dimensions, fine-grained variation, and time variation contribute to overall impression of voice quality in a speaker. This research takes the approach of conducting acoustic analysis on the same data as auditory-perceptual analysis, then rolling out this acoustic analysis to more data from a larger number of speakers. This will allow PPA to be validated and ensure that the differences in voice quality coded auditorily are present in the acoustic signal, increasing reliability. Furthermore, it will allow acoustic analysis to be interpreted in terms of auditory-perceptual descriptors of voice quality. Finally, expanding this acoustic analysis to a larger corpus allows a breadth of analysis that would not be possible using auditory-perceptual analysis alone.

However, carrying out an acoustic analysis of voice quality is not always straightforward. There are many potential measures, but interpreting them can be challenging. The way voice quality measurement is implemented often reveals researchers' ideas about what voice quality itself actually is, and will affect what is found, and overlooked, in a study on voice quality.

I turn now to the three types of acoustic measure that will be used in this research: Measures of harmonic spectral shape and slope, measures of inharmonic energy in the signal, and the use of f0 as a voice measure. Finally, I discuss other factors that need to be considered when implementing and interpreting acoustic measures.

4.2.2 Harmonic source spectral shape

When air is pushed through the glottis, setting the vocal folds (or indeed, the ventricular or aryepigottic folds) to vibrate and creating the source of voiced sounds, the configuration of the larynx during voicing affects how the folds vibrate. If the speech signal is inverse filtered by applying a filter to cancel out the vocal tract transfer function, removing the effect of formants, this allows us to examine the shape of the glottal

pulse to consider the way that airflow is modulated by the vocal folds. Examining the shape of the glottal source using the four-parameter Liljencrants-Fant model of glottal flow (Fant, Liljencrants, Lin, et al. 1985) reveals the effect of different laryngeal configurations on the resulting signal. Gobl & Chasaide (1992) examined how laryngeal voice quality affected the shape of the source pulse and resulting spectra by looking at different voice qualities produced by a phonetically trained speaker. They found that the high airflow and minimal adductive tension used to produce breathy voice results in features such as a symmetrical pulse and a long open phase, while tense voice shows a skewed glottal pulse, shorter open phase, and sharp glottal closures. Examining the glottal source in this way requires carefully controlled recording environment to maintain the phase response of the signal and hand-fitting the model to each glottal pulse in a very labour intensive process (Chasaide & Gobl 1993: 307). Because of this most acoustic measurement does not involve using inverse filtered signal; instead, it takes advantage of the fact that the shape of the glottal pulse in turn affects the acoustic signal, exciting particular frequencies in the spectrum. Considering the shape of the harmonic spectrum can therefore give insight into the underlying configuration of the vocal folds without having to use inverse filtering, which requires very high-quality recordings and can be time-consuming to conduct.

Gobl & Chasaide (1992) consider how the shape of the glottal pulse affects the relative amplitudes of formants, harmonics and higher-frequency parts of the spectrum. They find that in lax voice, harmonics in the region of the spectrum containing the first and second formant are attenuated relative to the first harmonic, and that this occurs to a greater degree for breathy voice, and an even greater degree for whispery voice. In averaged spectra, the amplitude is lower overall for lax voice than for modal, and even lower for breathy and whispery voice, and the spectrum is characterized by a dominance of the lower harmonics in all of these voice qualities. In whispery voice, however, they also find energy in the 4-5 kHz range, likely attributable to aperiodic noise from the turbulent airflow rather than increased harmonic energy in this area. By comparison, in tense voice and creaky voice the first harmonic is lower compared to the first formant, resulting from low airflow. The amplitude of the harmonics is higher overall in tense and creaky voice than in modal, and the longer closed phase of these voice qualities causes a reduction in amplitude with increasing frequency, levelling out the spectrum (Gobl & Chasaide 1992).

Because of the way that glottal constriction affects the harmonic spectrum, comparing the amplitude of different harmonics and different parts of the spectrum allows us to quantify the spectral slope, and in turn learn something about the degree of constriction in the larynx. A common measure that allows insight in to this is the difference between the first and second harmonic (H1-H2). Early acoustic analyses of voice quality (Bickley 1982, Fischer-Jørgensen 1967, Dave 1977) considered differences between phonemically modal and breathy vowels in languages such as Gujurati and

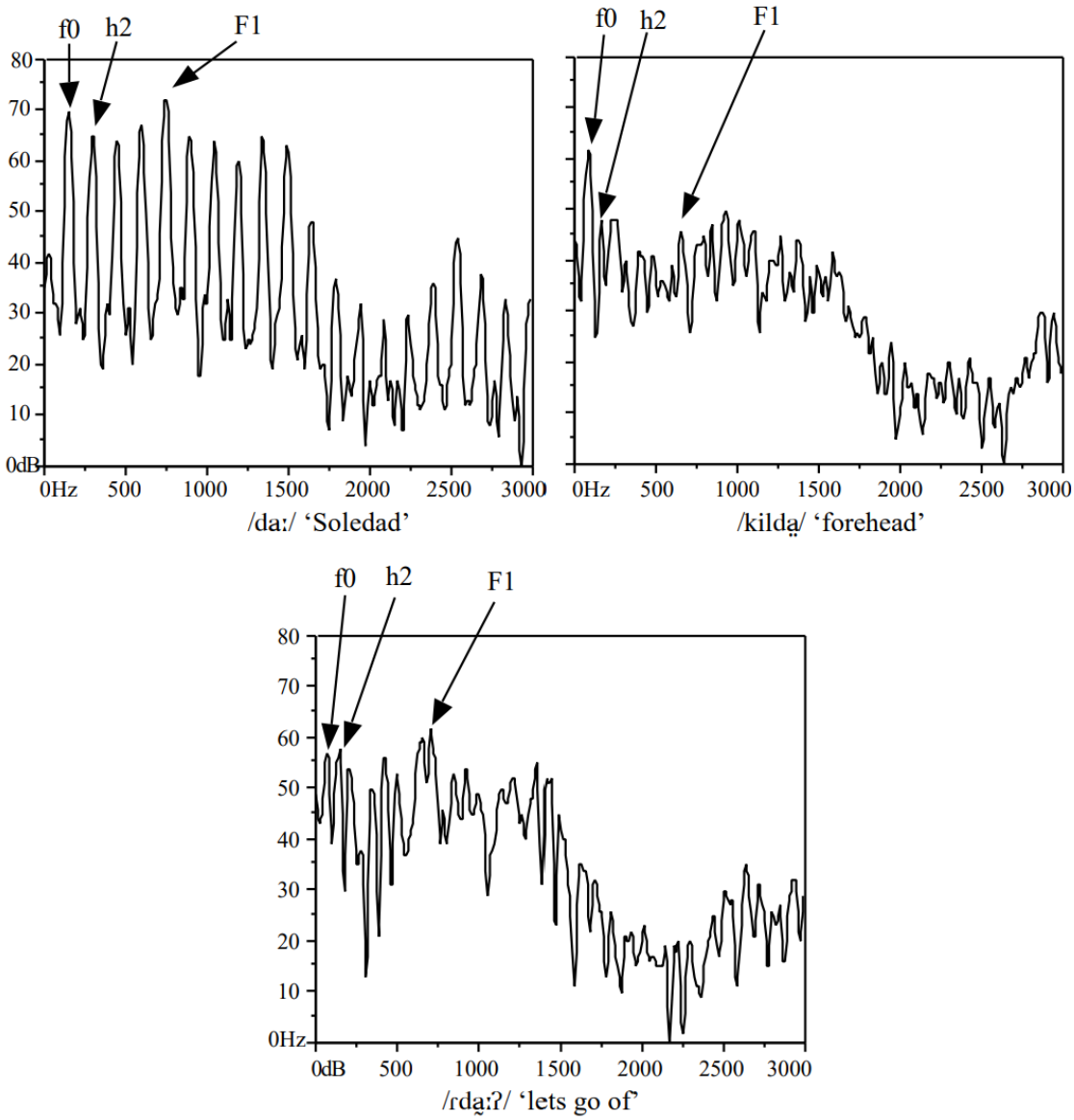


Figure 4.7: FFT spectra of modal, breathy, and creaky /a/ in the San Lucas Quiaviní Zapotec words /da:/ 'Soledad', /kilda̤/ 'forehead', and /rda̰:ʔ/ 'lets go of' (male speaker) from Gordon & Ladefoged (2001: 398), Figure 7.

!Xóõ, and found differences in the relative amplitude of these harmonics. H1-H2 differences have since been found in phonemic contrasts in other languages too: For example, as shown in Fig 4.7, in San Lucas Quiaviní Zapotec, the difference between H1 (marked as F0) and H2 (H1-H2) is higher and more positive for breathy voice and lower and more negative for creaky voice. As well as relating to overall degree of constriction, spectral tilt has also been linked 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 well as the medial thickness of the vocal folds and the amount of subglottal pressure (Zhang 2016a), which each in turn relate to the resulting quality of the sound. Since the development of formant-correction algorithms (Hanson 1997, Iseli & Alwan 2004, Iseli, Shue & Alwan 2007), harmonic measures are usually corrected for the influence of formants, with formant-correction being signified by an asterisk (e.g. H1*–H2*).

However, H1*–H2* is not the only measure of the harmonic source spectral shape. As Chai & Garellek (2019) note, the amplitude of H1 is not only higher or lower relative to H2, but in relation to the rest of the frequency spectrum, and the choice of H2 as a comparison point is arbitrary. Instead, Chai & Garellek (2019) propose the use of H1 alone as a measure of voice, using Root Mean Squared (RMS) energy to control for differences in Sound Pressure Level (SPL). Chai & Garellek (2019) compare how effective H1 is for distinguishing phonation types in !Xóõ and phrasal voice quality differences Mandarin, compared to H1*–H2*. They find that it makes clearer distinctions for these cases than H1*–H2*, with a higher H1 indicating breathy voice or increased spreading and a lower H1 indicating higher constriction or creaky voice.

Other measures of spectral tilt compare the energy of harmonics at different parts of the spectrum, or look at spectral tilt across the entire frequency range. This can be done in several ways, for example:

- Comparing two fixed harmonics other than just H1–H2, for example, H2–H4
- Comparing a fixed harmonic, such as H1, to the harmonic closest to a particular formant, such as F1, which can be noted as H1-A1, where An denotes the relevant formant
- Comparing a fixed harmonic to the amplitude of a harmonic closest to a fixed point in the frequency range, such as H4-2kHz
- Comparing the harmonics closest to two different fixed points in the frequency range, such as 2kHz-5kHz
- The general slope of the spectrum (termed 'Slope') and the tilt of the regression line through the spectrum (termed 'Tilt') (Maryn, De Bodt & Roy 2010)

As a generalisation, these different measures will have a similar relationship to constriction and spreading as H1–H2, higher values indicating breathy voice or increased spreading and lower values indicating higher constriction or creaky voice.

4.2.2.1 Combining multiple measures of spectral slope

Research that compares the phonetic realisation of phonation contrasts between different languages demonstrates that listeners may rely on acoustic information across the spectral slope to produce and perceive information about voice quality. Tian & Kuang (2021) compare how phonation contrasts are realised in four languages that include a phonation contrast with a 'breathier' and non-breathy phonation type: Gujurati, White Hmong, Southern Yi, and Shanghaiese. They consider H1*–H2*, H2*–H4*, H1*–A1*,

H1*–A2*, and H1*–A3*, and find that Shanghaiese is distinct from other languages in that there are not important differences in H1*–H2* between different phonation types, leading them to suggest that the 'breathier' phonation type in Shanghaiese is best described as whispery. Furthermore, while H1*–H2* does distinguish phonation types in the other three languages, the relative importance of each parameter in distinguishing the phonation types differs between different languages, with H1*–H2* being the most important in White Hmong and Southern Yi, but all other spectral slope measures being more important in Gujurati. Similarly, Keating et al. (2010) compare how phonation contrasts are realised in Gujarati (modal vs. breathy), Hmong (modal vs. breathy vs. creaky), and Yi (tense vs. lax). They find that while H1*–H2* distinguishes phonation contrasts across all four languages, other parameters (H1*–A1*, H1*–A2*, H1*–A3*) can also be used to distinguish phonation contrasts in Gujarati, Mazatec and Yi, but not in Hmong.

The value of combining measures can also be seen in research on the production and perception of voice quality. Starr (2015) looks at the realization of 'Sweet voice' in Japanese, a distinctive feminine vocal style used by voice actors. When compared to non-sweet voice performances by these voice actors, she finds sweet voice performances are characterised by higher values for H1–H2, H1–A1, and H1–A3, but lower values for 2k-4k. Furthermore, Bishop & Keating (2012) considered how listeners use voice quality to perceive speaker gender in isolated vowels, and found H1*–A3* affected listener perception of speaker gender, as did H2*–H4*, when voices were produced at low pitch, while H1*–H2* was less relevant. Together, this research demonstrates the value of using multiple measures of spectral slope to gain a more complete view of the shape of the spectrum and to evaluate which measures may be best suited to investigating voice quality in a particular context.

4.2.2.2 Spectral slope measures in the psychoacoustic model

Different research may use different measures from among these, using a measure or set of measures either because of convention or because of theoretically motivated reasons as to why it may be particularly informative for their research. However, in the psychoacoustic model, Kreiman et al. (2014) set out a selection of spectral slope measures which, when used together with the parameters in the model, aim to be able to model any voice quality, using only parameters to which listeners are sensitive. Kreiman, Gerratt & Antoñanzas-Barroso (2007) narrowed down a set of 78 measures by performing a principal components analysis of steady-state vowels produced by 70 speakers to establish which measures varied across speakers, then set out to establish a subset of these parameters that were both necessary and sufficient to model the quality of a voice. To do this, they used the UCLA voice synthesizer (Kreiman, Antoñanzas-Barroso & Gerratt 2010) to copy-synthesize hundreds of voices, then used 'same/different' tasks to see

if listeners could differentiate between the original voices and the voices that were copy synthesized using their set of parameters, refining these parameters to H1–H2, H2–H4, H4–2 kHz, 2 kHz–5 kHz. They then conducted further experiments that demonstrated that listeners were sensitive to these parameters and that each parameter was necessary in order to model voice quality, by varying each parameter in small steps and using 'same/different' tasks to establish the smallest change in each parameter that listeners could hear (Garellek et al. 2013, Kreiman & Gerratt 2012).

4.2.2.3 Acoustic and perceptual independence of spectral slope measures

Garellek et al. (2016) considers the set of spectral slope measures in the psychoacoustic model in more detail, considering that for the model to map the relationship between the acoustic signal and overall perceived quality as intended, the parameters within it should be independent of each other both acoustically and perceptually.

They first investigate whether the parameters are acoustically independent by copy-synthesizing 144 voice samples using the UCLA voice synthesizer and recording the values for H1–H2, H2–H4, H4–2 kHz, 2 kHz–5 kHz, as well as overall spectral roll-off, estimated to be approximately -12dB/octave by Flanagan (1957), to allow them to assess whether parameters were independent of this overall decrease in energy as frequencies increase. They found that model spectral slope parameters were predictable from one another, but that these relationships were due to the overall spectral roll-off.

They also investigated whether listeners could perceive differences in each individual parameter independently of the surrounding spectral shape. To do this, they created a synthetic copy of a voice and varied each parameter along a scale while the adjacent parameter was set to either a high or low value, and other parameters were held constant. They then conducted a 'same/different' experiment where listeners heard these stimuli and indicated whether they could hear a difference. They found that listeners could detect differences in H1–H2, H2–H4 and H4–2kHz independently of adjacent parameters, but that 2kHz–5kHz was more difficult to detect, and that listeners were less sensitive to differences in 2kHz–5kHz when H4–2kHz had a steep slope. They then conducted another experiment to consider listener sensitivity to assess whether listener sensitivity to this part of the spectrum depended on the amount of noise in the signal. They found that listeners had difficulty hearing differences in this 2kHz-5kHz range when there was a high amount of noise in the signal, and again found that listeners had difficulty hearing differences in this range when the slope at H4–2kHz was steep. However, if noise in the signal was low, and H4–2kHz was flat, listeners could detect differences in this high frequency part of the spectrum.

These measures are therefore acoustically and perceptually independent of each other, with the exception of 2kHz–5kHz, where perception depends on H4–2kHz and the

amount of noise in the signal. The results of these experiments therefore suggest that speakers and listeners may not only use H1–H2 to produce and perceive differences in voice quality, but also higher parts of the spectrum. Furthermore, because the different parameters defined in the model are largely independent of one another, differences in voice quality are likely best modelled using several of these parameters, rather than using a single parameter like H1–H2 as a proxy for the overall spectral shape.

In this research, I therefore draw on the insights gained from psychoacoustic research, and consider H1*–H2*, H2*–H4*, H4*–2kHz*, as combining measures will enable a more complete understanding of the voice than a single measure along. However, due to the fact that there is likely to be some level of background noise in the corpus data, I do not use 2kHz*–5kHz*, given that sensitivity to this region is context-dependent.

4.2.3 Noise measures

The configuration of the vocal folds and epilarynx during voicing can also lead to aperiodic noise in the signal, with modal voice containing more periodic energy and non-modal qualities containing comparatively more inharmonic energy. Noise can originate from aperiodicity in fundamental frequency, inefficient vocal fold closure, or epilaryngeal constriction. It can be investigated by a range of measures, but the most common ones are pitch and amplitude perturbation measures (jitter and shimmer), harmonics-to-noise ratio measures (HNR), and Cepstral Peak Prominence measures (CPP).

4.2.3.1 Jitter and shimmer

The use of perturbation measures originates in the work of Lieberman (1963), who found that disordered voices showed higher cycle-to-cycle variation in pitch, known as jitter. Jitter, along with shimmer, a measure of cycle-to-cycle variation in amplitude, has since been widely used, especially in clinical work. However, Kreiman & Gerratt (2005) noted that the role of jitter and shimmer in the perception of voice quality was not well established, and set out to establish the perceptual significance of jitter and shimmer relative to another measure, noise-to-signal ratio (NSR), a measure of the amount of aperiodic noise in the signal compared to the amount of periodic noise. They asked listeners to manipulate jitter, shimmer, and NSR in synthetic stimuli, until the synthetic voices matched the target voices, and found that listeners did not agree on how much jitter and shimmer should be added to the synthetic voices to replicate target stimuli, and were unable to separate the effect of each parameter.

4.2.3.2 Harmonics-to-noise ratio

However, listeners in this experiment did agree on their ratings of NSR, a parameter which belongs to a family of harmonics-to-noise or signal-to-noise ratio measurements that aim to compare the ratio of harmonic and inharmonic noise. This technique was initially developed by Kojima et al. (1980), who observed that, in sustained vowels produced by speakers with voice disorders, the clarity of the harmonics in a narrowband spectrogram was obscured by noise components, and developed a technique to separate out the harmonic components from the noise components and compare them. This technique was later simplified by Yumoto, Gould & Baer (1982), who did this by averaging the energy of the waveform over 50 consecutive pitch periods in a sustained vowel and comparing this to the averaged energy of the differences between individual periods.

As de Krom (1993) notes, other techniques for estimating harmonics-to-noise ratio also existed at this time, but these varied between different studies and precise methods were not always made explicit. de Krom (1993) therefore aimed to standardise HNR measurements through developing a new technique for estimating HNR. This differs from previous methods in that it uses the log power spectrum, known as the cepstrum[1], to discriminate between harmonic energy and noise energy. de Krom (1993) takes advantage of how periodic energy originating from the fundamental and the harmonics occupies fairly distinct 'quefrencies' (log-transformed frequencies) from aperiodic energy that comes from noise in the cepstrum. This allows a comb-lifter (filter) to be applied to the cepstrum, and lifter the rahmonic (log-transformed harmonic) energy out of the cepstrum. This allows the harmonic and inharmonic parts of the spectrum to be estimated and compared to one another. de Krom (1993) argues that as this does not need to be conducted on a long, stationary signal, it has the advantage of being usable on connected speech.

Boersma (1993) has also implemented HNR in Praat, working in the autocorrelation domain. Boersma (1993) argues this approach is less dependent on window length and period and is more resistant to rapidly changing sounds than previous approaches.

Cepstral Peak Prominence (CPP) is also comes under the umbrella of Harmonics-to-Noise Ratio measures. First developed by Hillenbrand & Houde (1996), CPP also takes advantage the manifestation of harmonic energy in the cepstrum. As shown in Fig 4.8 (top right), the cepstrum of a non-breathy voiced signal shows a more prominent cepstral peak corresponding to the regularity of harmonics. In less periodic, less modal signals (Fig 4.8, bottom right), the prominence of this peak is reduced. To calculate the prominence of this peak, Hillenbrand, Cleveland & Erickson (1994) fit a regression

[1]Terminology related to the cepstral domain involves rearranging letters to refer to analogous concepts, following Bogert (1963).

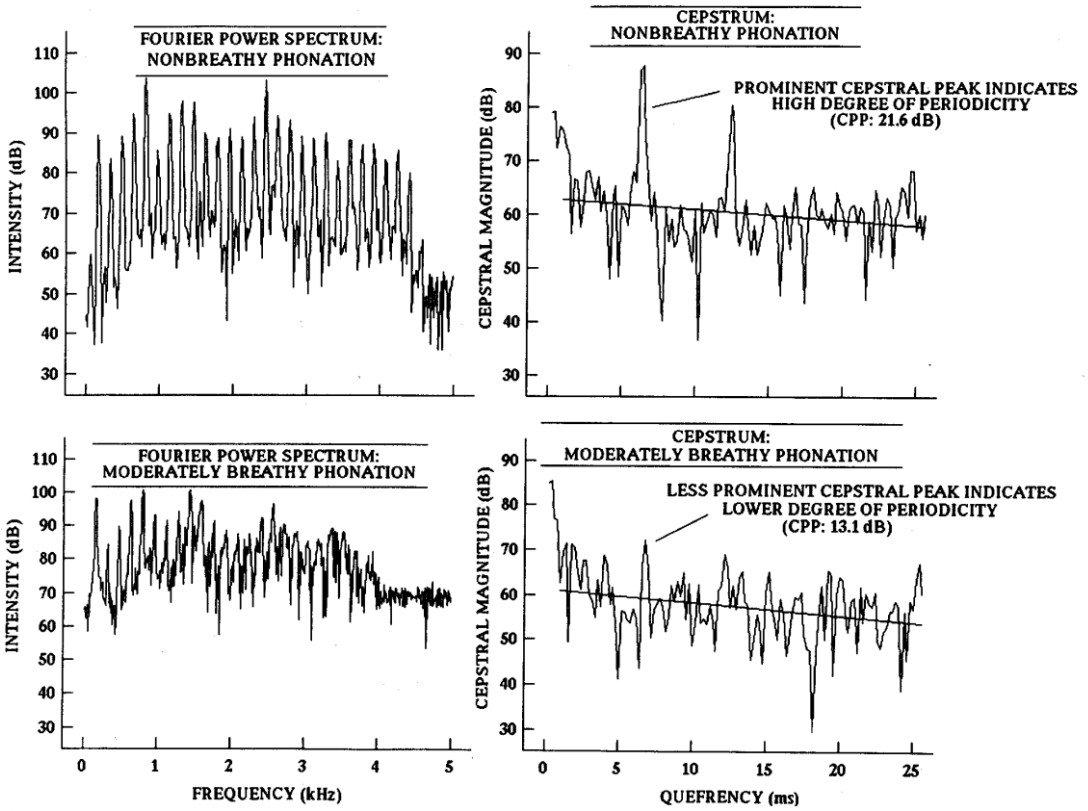


FIGURE 2. Fourier power spectra and cepstral representations for nonbreathy and moderately breathy signals. The nonbreathy signal received a rating of 0.8 on a 0–10 rating scale, and the moderately breathy signal received a rating of 4.0. The linear regression line relating quefrency to cepstral magnitude is used in the CPP measure to normalize the cepstral peak for overall amplitude. CPP is the difference in amplitude between the cepstral peak and the value on the regression line that is directly below the peak.

Figure 4.8: An illustration of cepstral peak prominence from Hillenbrand & Houde (1996)

line to the cepstrum, then calculate the amplitude difference between the cepstral peak and the value of the regression line at that quefrency.

4.2.3.3 Perception of noise measures

Capturing inharmonic spectral noise also forms part of Kreiman et al.'s 2014 psychoacoustic model of voice quality. Early iterations of the model recognised a need to model inharmonic spectral noise correctly. For example, Kreiman, Gerratt & Antoñanzas-Barroso (2007) found that high-frequency noise excitation varied between voices and Kreiman & Gerratt (2005) considered whether jitter and shimmer were perceptible independent of overall harmonics-to-noise ratio. Later, Signorello et al. (2016) considered variation in the shape of the inharmonic spectral source, proposing that a four-piece model of the noise spectrum accounts for most of the variance in inharmonic spectral shape. However, as Keating et al. (2021: 8) note, it is still unclear how sensitive listeners are to inharmonic spectral shape and there is no consensus as to how the shape of the noise spectrum should be modelled.

Fine-grained acoustic difference in the noise spectrum may help to differentiate different phonation types. While HNR and CPP look at the overall proportion of noise in the spectrum, this noise can originate from two main sources: aperiodicity in vibration at the source, as in the case of creaky voice, or turbulent noise, as in the case of whispery voice. In gross HNR measures, these two types of noise cannot be differentiated, but specific measures of HNR that look at specific frequency bands may help to differentiate these more clearly. De Krom (1995) looks at HNR in different frequency bands in breathy and rough vowels and found that low HNR in frequency bands up to 2kHz was typical of roughness, and the emergence of spectral noise from 2-5kHz was more typical of breathiness, though there was not a clear-cut distinction between breathiness and roughness. Keating, Garellek & Kreiman (2015) looked at HNR in different frequency bands in creaky voice and speculate that the 0-500 Hz HNR band (HNR05) may be the most sensitive to the irregular fundamental frequency of some types of creaky voice, presumably because the fundamental frequency of most speech lies in this frequency range and so the f0 irregularity will contribute most to decreased HNR in this range. This is supported by Keating et al. (2021), who compare the phonetic realisation of phonation contrasts in 11 languages and find that HNR05 contributes to differentiating creaky and modal voice.

4.2.4 Fundamental frequency as a measure of voice quality

Fundamental frequency is perhaps not traditionally seen as a measure of voice quality, but differences in f0 do characterise certain phonation types, such as creak and falsetto. Relationships between the laryngeal mechanisms responsible for raising and lowering pitch and those responsible for creating changes in phonation can also mean that certain phonation types tend to favour certain pitch ranges; Laver (1980) notes breathy voices tend to be low in pitch. Furthermore, measurements for most other voice quality measures rely on accurate measurements for f0, because the first harmonic is the fundamental and higher harmonics are multiples of this value. Accurate measurements of f0 are therefore essential for acoustic analysis of voice.

There are many f0 trackers available, with Tsanas et al. (2014) comparing 10 widely used f0 trackers, but noting the existence of at least 70. Tsanas et al. (2014) outline how these f0 trackers differ in many ways, with a major division being between those that work in the time-domain (e.g. Boersma 1993), or in the frequency domain (e.g. Sun 2002). The number of available f0 trackers may reflect the difficulty of creating an algorithm that functions across diverse f0 levels, voice qualities and recording contexts. Some of the main difficulties, summarised by Talkin (1995), include:

- f0 being a time-varying aspect of the signal and changing between cycles

- Subharmonic component frequencies that exist as fractions of the fundamental frequency

- Harmonic frequencies, which exist as multiples of the fundamental frequency

- Differentiating the fundamental from periodic background noise

As Tsanas et al. (2014: 2889) note, there may be no single f0 tracking algorithm that is appropriate for all purposes; rather, some trackers may perform better in some circumstances and with certain types of voices, but worse for others. Tsanas et al. (2014) compare the performance of 10 f0 tracking algorithms on isolated vowels both produced naturally and created using a synthesizer, comparing the outputs of each f0-tracking algorithm with the EGG-tracked f0 or intended synthetic f0 respectively. They found that time-domain approaches generally performed worse than frequency-domain approaches. Different algorithms had different strategies and degrees of success for considering the challenges of robust f0 tracking, with some developing successful strategies for differentiating subharmonics from the fundamental (e.g. Camacho & Harris 2008), and some being particularly robust to background noise (e.g. Boersma 1993).

When tracking f0 in voice quality research, understanding the advantages and limitations of a particular f0 tracker and balancing these with practical considerations is perhaps more important than attempting to find one with no drawbacks. For example, previous versions of VoiceSauce (Shue et al. 2011), a voice analysis program, used the Nearly Defect-Free (NDF) algorithm Hideki Kawahara et al. (2005) as the default f0 tracker. However, more recent versions instead use the XSX algorithm Kawahara et al. (2008), an updated algorithm that performed less successfully overall in Tsanas et al.'s 2014 analysis. However, NDF was known to crash on some signals (Vicenik et al. 2022) and was very computationally demanding (Kawahara et al. 2008), so this trade-off in terms of overall performance is offset by increased speed and practicality.

4.2.5 Considerations for approaching acoustic analysis

Beyond the exact measurement used, there are other things to consider when analysing voice quality acoustically. Here I consider how different implementations of acoustic measurement reflect diverse understandings of voice quality, how acoustic measurement can capture different dimensions of the voice, and the influence of segmental context and recording environment.

4.2.5.1 Unit of analysis

The measures described thus far can be implemented on a range of time spans that are appropriate for different research focuses. While spectral measures are usually conducted on vowels, research that follows Laver (1980) in conceiving of voice quality as a long-term setting may take the approach of averaging spectra over a longer stretch of speech for a speaker. In this case, the Long-Term Average Spectrum (LTAS) can be conducted on either read speech using a phonetically-balanced passage to control for the effects of segemental features (Löfqvist 1986: 471), or across at least 30 seconds of continuous speech that is not phonetically balanced, after which point the effects of segmental features begin to average out (Li, Hughes & House 1969).

If voice quality is seen as a long-term feature, a long-term frequency-based acoustic measurement can be argued to be more appropriate than short-term measures (Pittam & Gallois 1986: 2). The analysis can be conducted only across voiced segments, as in Löfqvist (1986), to reveal information about laryngeal voice quality, or across all segments, as in Nolan (1983), which may reveal information about supra-laryngeal voice quality too. The idea behind conducting an analysis of voice quality in this way is that any short-term variation in voice quality will be averaged out, reveal something about the speaker's overall voice quality.

Some examples of LTAS for different laryngeal voice qualities, produced by Nolan (1983), are shown in Fig 4.9. The LTAS can be examined visually, but as with more short-term spectra, it is also possible to compare the relative amount of energy in different frequency regions of the LTAS.

Nolan (1983) considers differences in the LTAS between different voice qualities, both laryngeal and supra-laryngeal, produced by himself and John Laver. To investigate differences, he compares the upper part of the spectrum, between 1.5-3 kHz (shown in red in Fig 4.9), to the lower part of the spectrum, between 0-1.5 kHz (shown in blue in Fig 4.9), as well as an approximation of the dB/octave slope of the spectrum. Although neither of these measures fully discriminated the different voice quality settings, there were tendencies for different laryngeal settings to produce certain characteristics in the LTAS: Qualities with falsetto or whispery components appeared to have steeper spectral slopes, and qualities with creak and harsh components appeared to have more level spectra. Work by Pittam (1985, 1987) also found that different breathy, whispery, creaky and tense voice can be differentiated through characteristics of the LTAS, looking instead at intervals of 200Hz, 1.5 Bark, and 200 mel from 0-2kHz, and finding that 1.5 Bark and 200 mel intervals were more successful than Hertz intervals for differentiating different voice qualities.

However, Nolan (1983) and Pittam (1985, 1987) both examined the LTAS of voice

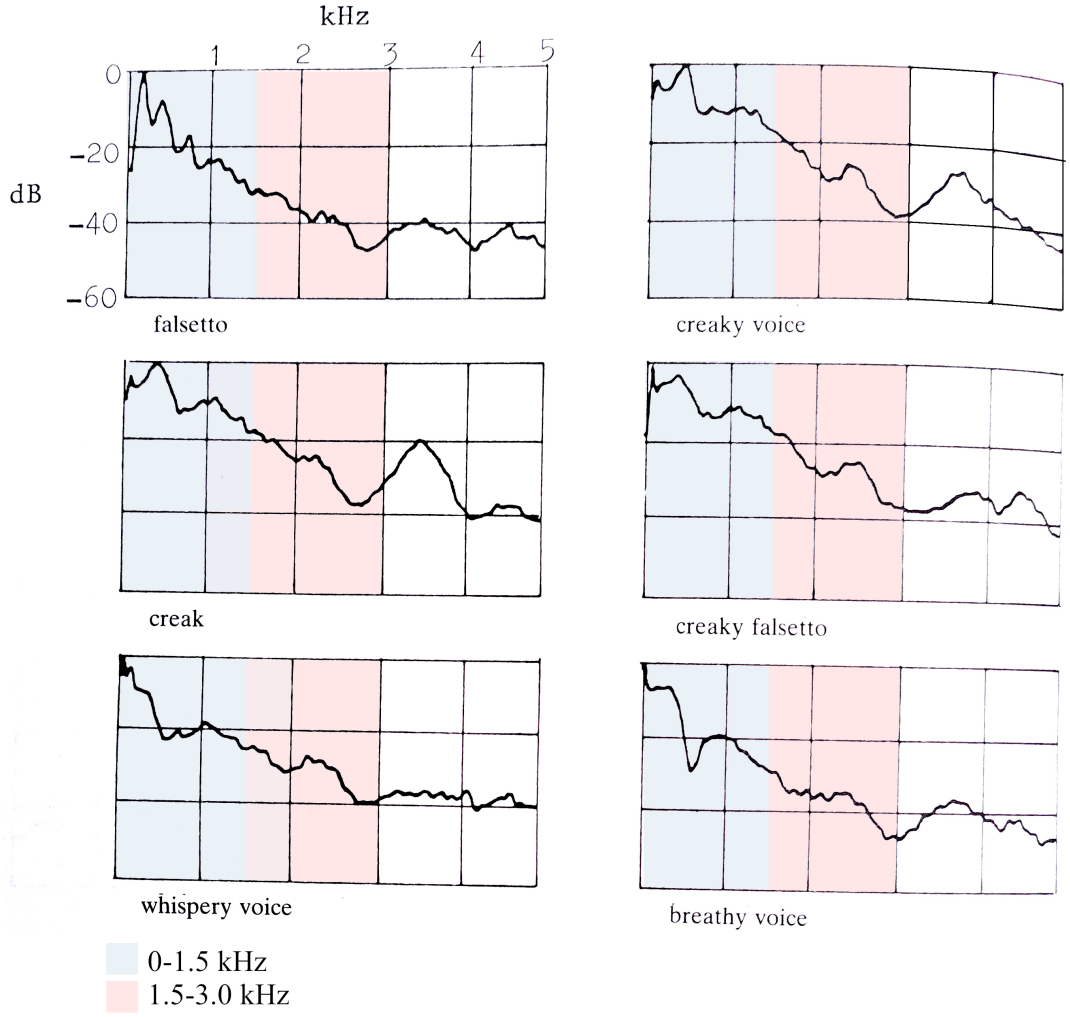


Figure 4.9: Long-Term Average Spectra of falsetto, creak, creaky voice, creaky falsetto, whispery voice, and breathy voice from Nolan (1983). Own annotation shows the 0-1.5 Hz range in blue and 1.5-3 Hz range in red.

samples which had been intentionally produced with different voice quality settings by phonetically-trained speakers. In cases where stimuli are not elicited with a particular voice quality, LTAS may be less useful. For example, Löfqvist (1986) looked at whether LTAS could be of use in clinical settings to differentiate between voices with and without voice disorders, between different types of voice disorder, and for tracking changes in a patient's voice over time. Löfqvist (1986) finds a limitation of the LTAS in that it does not clearly distinguish between harmonic and inharmonic components of the spectrum. One speaker that Löfqvist (1986) considers has bilateral vocal fold paralysis and produces a very breathy voice. The LTAS for this speaker's voice shows a high amount of energy between 1-5kHz, corresponding to noise generated by incomplete vocal fold closure. However, in the absence of an auditory impression of the voice, energy in the part of the spectrum could have been attributed to harmonic components of the voice source.

Löfqvist (1986) also highlights that the LTAS may be limited because of the amount of variability within and between voices. With an LTAS, short-term variation is averaged out across an entire sample from a speaker. Because of this, the LTAS is likely limited in usefulness in the case of the present study, where within-speaker variation and variability in voice quality is taken as an important part of what characterises long-term voice quality.

4.2.5.2 Automatic detection of creaky voice

As covered in Section 4.1, sociolinguistic studies of voice quality that take an auditory-perceptual approach consider that each vowel or syllable can be coded for the presence or absence of a particular voice quality, capturing within-speaker variation. This is particularly common for analysing creak and automated acoustic approaches have been developed that mirror this, automatically the presence of creaky voice.

Automatic coding of creak has the advantage of comparability with studies conducted using auditory coding of creaky voice, and is intuitive to interpret. Furthermore, I propose that separating creaky and non-creaky speech as a first step may help us to interpret other acoustic measures. A certain value for H1*–H2*, for example, can be very difficult to interpret in isolation. However, if creaky voiced speech has already been detected automatically, low values of H1*–H2* in non-creaky voicing can be interpreted as tense voice, while high values of H1*–H2* in non-creaky voice suggest to breathy voice, and high values of H1*–H2* in creaky voice suggest non-constricted creak. Automatic creaky voice detection therefore allows more straightforward interpretation of other measures.

Several automatic creaky voice detection methods have been developed. Ishi et al. (2008) developed an automatic creaky voice detection method that relied on the aperi-

odicity of creaky voice. This method identifies potential glottal pulses, then examines the periodicity of pulses to decide whether or not they are potentially creaky, and then looks at the similarity between successive pulses to detect stretches of creaky voicing. The resulting algorithm correctly identified 74% of manually-coded creaky voice. 13% of creaky voice identified by the algorithm was insertion error, consisting of 3% falsely identified creaky voice in voiced segments, and 10% in voiceless segments.

Kane, Drugman & Gobl (2013) propose an alternative, Resonator-Based Creaky Voice (RBCV) detection technique, which involves detecting creaky voice using a neural network. The neural network uses a combination of two acoustic measures that are taken after a resonator is applied to the LP-residual signal to make decisions about the presence or absence of creaky voice. According to Kane, Drugman & Gobl (2013), RBCV identifies creaky voice in 81% of cases, with 27% of creaky voice identified consisting of false alarms. In a direct comparison to the method developed by Ishi et al. (2008), they find that RBCV performs better than Ishi et al. (2008), which they found flagged a high number of false positives in speakers with low non-creaky f0.

Dorreen (2017) and Dallaston & Docherty (2019) propose an f0-based approach to detecting creaky voice, which rests on the idea that creak is often characterised by low f0 (Keating, Garellek & Kreiman 2015) relative to the rest of a speaker's f0 range. Dallaston & Docherty (2019) argue that with an f0 tracker that is robust at low f0s, it should be possible to distinguish between creaky and non-creaky speech using a speaker-specific antimode; that is, the location of the f0 value that occurs least often between the modes of creaky and non-creaky speech.

Dallaston & Docherty (2019) describe and evaluate the effectiveness of this approach using REAPER (Talkin 2015), an f0 tracker which Dorreen (2017) found was able to detect Glottal Closure Instants (GCIs) reliably at low f0 and allows calculation of f0 for each individual glottal cycle. Dallaston & Docherty (2019) use REAPER to implement this f0-based automatic detection of creaky voice in a corpus of read and re-told speech produced by 42 speakers aged 18-40. They identify the f0 antimode for each speaker and use this to identify creak and calculate the overall percentage of creaky voice used by each speaker.

Upon hand-checking a random sample of 15% of the speech that REAPER annotated as creaky and as non-creaky, they found that 81% of what REAPER annotated as creaky was accurately annotated, and that 98% of what REAPER annotated as not creaky was accurately annotated, with similar performance for male and female speakers. Dallaston & Docherty (2019: 535) conclude that this approach is effective at achieving 'coarse-grained' estimates of creaky voice prevalence across and between speakers and is accurate enough to be used on large datasets, but stress that this approach will not always agree with an auditory analysis of creak because it relies on a

single acoustic cue.

In this research, I follow Dallaston & Docherty (2019) in automatically detecting creaky voice using an f0-based method, but build on this by following automatic identification of creak with further acoustic analysis of non-creaky speech.

4.2.6 Capturing different dimensions on voice quality

4.2.6.1 Different dimensions in a single moment

Acoustic analysis has the potential to allow consideration of how different dimensions combine to contribute to the overall impression of a speaker's voice. However, doing this requires careful consideration of how different acoustic measures connect to different dimensions of voice quality, and how a difference in a single acoustic measurement could reflect any one of several different dimensions of voice quality, which is not always straightforward.

Any voice sample is made up of many different frequencies, both harmonic and inharmonic, which combine to produce its auditory quality; at any given moment, a single acoustic measurement can only capture some of this information. As Garellek (2019: 89) explores, researchers often connect lower values of H1–H2 to creaky voice, and higher values to breathiness. Garellek (2019: 89) notes that raw spectral tilt values do not index a precise quality, and are relative within the voice of each speaker. Because of this, if we have measured H1-H2 in two vowels and find that H1-H2 is higher in one than the other, it is impossible to know from H1-H2 alone if this difference is because one is more creaky and the other more breathy, if one if less breathy and the other is more breathy, or if one is more creaky and the other is close to modal voice.

However, differences in H1-H2 become easier to interpret when used alongside noise measures. Garellek (2019) illustrates the way that H1-H2 is clearer to interpret when used alongside HNR, where higher values indicate more aperiodicity in the signal. For example, if H1-H2 is higher in Vowel A than in Vowel B, HNR may show that Vowel A has less noise than Vowel B and suggest that Vowel A is closer to modal while vowel B is more creaky. However, if vowel A has more noise than vowel B, this suggests that Vowel A may be breathier, while Vowel B may be creaky.

The use of combining spectral tilt and noise measures is also illustrated in the case of Tian & Kuang's (2021) study of how phonation contrasts are realised in Gujurati, White Hmong, Southern Yi, and Shanghaiese. In addition to the previously discussed finding that certain spectral tilt measures differentiated phonation contrasts better in some languages than in others, they also considered the role of noise measures. In Shanghaiese, CPP, HNR05, HNR15, HNR25 and HNR35 were the most important cues

for differentiating the two phonation types, suggesting a whispery-modal contrast. In Southern Yi, these noise cues had very minimal influence, and spectral tilt measures were much more important, suggesting a tense-lax contrast. In White Hmong and Gujurati, spectral tilt measures were more important than noise measures, but noise measures still played a role, suggesting a breathy-modal contrast. Examining differences in both noise and tilt measures therefore allows Tian & Kuang (2021) to consider the realisation of each of these 'breathier' phonation types than using either noise or tilt measures alone.

Furthermore, these dimensions of voice quality cannot be separated from each other because spectral shape and noise measures also interact in perception. As Kreiman & Gerratt (2012) find in a series of Just-Noticable-Difference experiments, there is a complex perceptual relationship between Noise-to-Harmonics Ratio (NHR) and spectral slope. Listeners were more sensitive to small differences in NHR when NHR was high, but this decreased when spectral slope was steep. Additionally, listeners were less sensitive to changes in spectral slope when there was a steeper spectral roll-off, and this effect also depended on the level of noise present. Furthermore, it was not just the amount of noise that affected listeners' ability to perceive small differences in noise and spectral slope, but also the shape of the noise spectrum: For example, listeners were more sensitive to changes in slope and the amount of noise when the noise spectrum was flat. These results suggest that it is difficult to separate noise and harmonic energy as a listener, and in turn this implies that combining noise and spectral slope measurements is useful for analysing the voice.

4.2.6.2 Different dimensions at different time points

Szakay & Torgersen's (2015, 2019) examination of voice quality in London English is a case study which demonstrates the difficulty of capturing both categorical and continuous dimensions of voice quality across time. They consider whether a particular voice quality can be associated with Multicutural London English (MLE), a variety typically spoken by young speakers in areas of London with high numbers of recent immigrants like Hackney. Szakay & Torgersen (2015) use H1-H2 and f0 to compare voice quality of speakers in Hackney, looking at both speakers with local family roots (Anglo) and speakers who were children or grandchildren of immigrants (non-Anglo), with voice quality in Anglo speakers in Havering in outer London.

Szakay & Torgersen (2015) reported patterns in H1-H2 that suggested that voice quality was a feature of MLE and varied between Hackney and Havering. This included low H1-H2 for Anglo female speakers in Hackney, suggesting creaky voice, but high H1-H2 for Anglo female speakers in Havering, suggesting breathy voice. Revisiting this data in Szakay & Torgersen (2019), however, they found different results using

Dallaston & Docherty's 2019 f0-based method of automatically identifying creak. Anglo female speakers from Hackney, who had been the creakiest according to the H1-H2 method, used less creak than any other group when this was defined as f0 dropping below that speaker's antimode. Furthermore, Anglo female speakers from Havering, who had been the breathiest speakers in terms of H1-H2, actually used more low f0 creak than any other group.

Szakay & Torgersen (2019) attribute these different results to the increased precision of the f0 tracker used, REAPER, at low f0 values, when compared to Praat's f0 tracker used in Szakay & Torgersen (2019). This explains the results in part: if Praat had difficulties detecting f0 at low frequencies, then measurements for H1-H2 would be either incorrect or not possible for much of the creaky speech. However, this difference also reflects a shift from conceptualising variation in voice quality as continuous, between more-creaky lower H1-H2 values and more-breathy higher H1-H2 values, to an approach where creaky voice is seen as distinct from non-creaky voice. As Fig 4.10, from Szakay & Torgersen (2015), shows, a bimodal distribution of f0 values was already apparent for these speakers when f0 was measured in Praat (albeit with fewer values being picked up in this range), with female speakers exhibiting a cluster of higher-f0, higher-H1-H2, non-creaky speech as well as a cluster of lower-f0, lower-H1-H2, creaky speech. Taking both analyses together suggests speakers may be characterised by using high rates of creaky voice *and* a high degree of breathiness in their non-creaky speech. If voice quality is approached solely as a continuous creaky-breathy scale, or solely categorical phenomenon where creak is present or absent, the voice of a speaker who is both more breathy and uses more creak is impossible to measure.

Figure 3: The effect of f0 on H1-H2 by speaker groups.

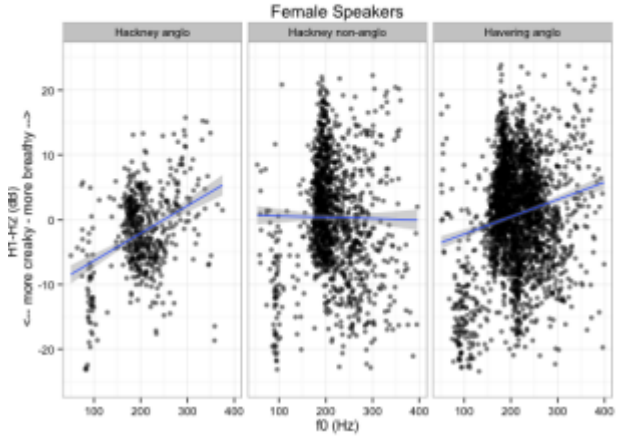


Figure 4.10: Figure 3 from Szakay & Torgersen (2015: 3), showing a bimodal distribution of f0 for female speakers

4.2.7 The influence of context and other parts of the acoustic signal

Much research on the acoustics of voice focuses on steady-state vowels recorded in ideal environments. However, the present study uses corpus data that includes a degree of background noise, and conducts measurements spontaneous speech, which may be influenced by phonetic context. Here, I discuss the influence of these factors on acoustic measurement.

4.2.7.1 Recording environment and microphone

Bottalico et al. (2018), Deliyski, Evans & Shaw (2005), Deliyski, Shaw & Evans (2005), van der Woerd et al. (2020) compare the performance of various acoustic voice measures and find that recording environment and microphone set-up can affect acoustic measurement of voice parameters, with background noise, reverberation in the room, microphone specification, and microphone distance from the mouth all having an effect.

Bottalico et al. (2018) compare acoustic measures collected in a sound booth, ensuring low background noise and short reverberation time, to three other environments, each with increased levels of background noise and reverberation time. They hypothesised that frequency-based measures like jitter and CPPS would be less affected by the environment than amplitude-based measures like shimmer and HNR, which might more sensitive to the filtering effects that come from increased reverberation in a room. They do find that jitter and CPPS are more independent of the acoustic characteristics of the room, followed by shimmer, HNR, pitch strength (a measure of perceived pitch), AVQI, and spectrum tilt, and conclude that the effect of recording environment is stronger than that of microphone.

van der Woerd et al. (2020) also consider the impact of recording environment, and compare measures of f0, jitter, shimmer, HNR and CPP from recordings taken in a sound-proof room and in a quiet office environment that is not sound treated. They find significant differences in shimmer, HNR and CPP between the two recording environments, but suggest that the differences between the two settings are still small, particularly for CPP, and may not have a clinical significance when it comes to interpretation of measures.

Microphone specification and set-up can also affect acoustic measures. Švec & Granqvist (2010) review technical and voice literature to produce a set of guidelines for selecting microphones for voice research. They firstly cover recommendations concerning frequency response, and suggest that microphones used should first of all have a flat frequency response curve, so that the microphone used is equally sensitive to

different frequencies, as well as a frequency response range that is capable of capturing the lowest and highest frequencies produced by human voices and perceptible to the human ear, going down to at most 50 Hz, but ideally to 10 Hz, and extending upwards to at least 8000 Hz, but depending on the exact frequencies of interest, potentially to 20,000 Hz. Inverse filtering requires additional specifications, such as a flat phase response to preserve the exact shape of the waveform.

Švec & Granqvist (2010) then discuss directionality of microphones. Microphone directionality is commonly either omnidirectional, so that the microphone is equally sensitive to sounds coming from all directions, or cardioid, where the microphone picks up the signal coming from the front well but suppresses noise coming from the back, which helps to suppress background noise from the room. Suppressing background noise is advantageous for voice quality measurement, but cardioid microphones also have the disadvantage of creating a proximity effect, where lower frequencies are boosted when the microphone is closer to the mouth, affecting the flatness of the frequency response. Švec & Granqvist (2010) note that some microphones will specify a particular distance at which they have a flat frequency response curve, and recommend using cardioid microphones only at this distance, and to avoid using cardiod microphones where this distance is not known or specified.

Bottalico et al. (2018) considered how acoustic measures were affected by directionality, frequency response and distance in practice, and found that while the microphone specification was important, with HNR in particular varying between different microphones, it was less so than the acoustics of the room. They find that head-mounted microphones perform poorly in comparison to other microphones in their research, which they suggest could be a result of the proximity effect. However, one of the head-mounted microphones considered had a limited frequency response of 100-10,000 Hz, and the other was an omnidirectional head-mounted microphone which would not have been affected by the proximity effect, so it is unclear whether this was in fact a result of the proximity effect.

4.2.7.2 Vowel formants

The presence of a formant boosts the harmonics around it, and in turn this means that formants can influence any measures that rely on measurements of harmonics. For example, in high vowels, F1 influences the amplitude of the lower harmonics: In early studies of voice quality such as Henton & Bladon (1985), this led to only open vowels being considered in the analysis, because in these vowels F1 was high enough not to interfere with measurements of H1 and H2. The influence of formants became apparent in Henton & Bladon's (1988) analysis when they examined [ɒ], which was closer than the other vowels examined, [ɑ], [æ], and [ʌ], and was in turn affected by the

first formant, and displayed a lower amplitude difference between the first and second harmonic.

However, Hanson (1997) developed a formula for correcting the effects of first formant on harmonic measurements. Iseli & Alwan (2004) and Iseli, Shue & Alwan (2007) later developed this formula to account for the effect of higher formants and formant bandwidths as well. This formula has been incorporated into software for taking voice quality measurements, such as VoiceSauce (Shue et al. 2011), minimizing the effect of formants on harmonic measurements in more recent research. However, this means that formant-corrected harmonic amplitude measurements are dependent on accurate formant measurements, and if formants are tracked incorrectly this will lead to errors in the harmonic measurements.

4.2.7.3 Nasality

Simpson (2012) argues that the difference between the first and second harmonic should not be used to compare breathiness between male and female speakers. He outlines the theoretical reason behind this: In the presence of nasality, the first nasal pole occurs at approximately 200-300 Hz. If a speaker has a fundamental frequency of 200-300 Hz, as may be the case for many female speakers, this nasal pole will occur in the region of the fundamental frequency, and thus enhance this first harmonic. However, in the case that a speaker has a lower fundamental frequency, for example, 120 Hz, as is more typical of male speakers, this nasal pole is expressed as an enhancement of the second and third harmonics. A comparison of H1-H2 between male and female speakers, Simpson argues, may be influenced by any nasality that may manifest as an enhanced H1 in female speakers but as an enhanced H2 in male speakers.

Nasality can also affect noise measurements. Madill et al. (2019) investigated how Praat's implementation of HNR and the Analysis of Dysphonia in Speech and Voice implementation of CPP differed between vowels, nasalized vowels, and nasal consonants, hypothesizing the increased nasality would decrease CPP, but not affect HNR. They found that CPP did decrease in the presence of nasality, and suggested that this was related to nasal coupling, which reduces the intensity of the signal. They also found that HNR increased in the presence of nasality, but this effect was smaller than the effect of nasality on CPP. They suggested that increased HNR may be related to increased periodicity in nasal sounds, rather than to an nasality affecting measurements of HNR.

4.2.8 Summary

In this section, I have given an overview of different approaches to acoustic analysis and considered how different dimensions of voice quality can be taken into account. In this thesis, I take the view that a multi-pronged approach might be appropriate for capturing the multi-dimensionality of voice quality within a single time point, and across different time points. I therefore attempt to bring together two different approaches: Automatic f0-based detection of creaky voice, following Dorreen (2017), Dallaston & Docherty (2019) and Szakay & Torgersen (2019), and multi-measure acoustic analysis that involves both noise and spectral tilt measures, taking some of the key principles of the psychoacoustic model into account (Kreiman et al. 2021). In doing so, I hope to allow clearer interpretation of acoustic measurement of voice quality: By first separating creaky voice quality, I reduce the amount of possible phonation types that remain. This then allows spectral tilt measures to be more clearly interpreted in terms of glottal constriction, and noise measures to be interpreted in terms of additional aperiodic noise, rather than f0 irregularity.

Part II

1st order indexicality: A linked auditory-perceptual and acoustic study of Scottish voice quality

Chapter 5

Introduction

In this part of the thesis, I asked how laryngeal voice quality varies in Scots varieties by gender, age and area. Specifically, I used data from SCOSYA (Smith et al. 2019) to examine differences between speakers of two Central Scots varieties, Glasgow and Lothian, as well speakers of Insular Scots. I looked at male and female speakers from each region, as well as older speakers (approximately 65+) and younger speakers (approximately 18-25).

To investigate this question, I began with a linked auditory-perceptual and acoustic analysis of voice quality in 90 seconds of speech from 24 speakers (eight from each of Glasgow, Lothian and Shetland, with two speakers in each cell for each age/gender/area group). I aimed here to link acoustic measurement and automatic f0-based identification of creak to auditory-perceptual labels for voices, so that further acoustic analysis could be interpreted with reference to previous research on voice quality in Scotland that uses VPA. I also aimed to assess the feasibility of conducting acoustic analysis on spontaneous speech collected outside of a lab setting, as previous work using similar approaches has tended to use high-quality lab recordings. Chapter 6 presents an auditory-perceptual analysis of these speakers conducted using PPA. Chapter 7 aims to link PPA coded creak to automatically detected creak using an f0-based method, then Chapter 8 links variation in non-creaky voice quality coded with PPA to acoustic analysis of voice quality with H1*–H2*, H2*–H4*, H4*-2kHz* and CPP.

Building on the approach taken with the smaller subset of speakers, I then considered how voice quality varied in a larger sample of 95 speakers from SCOSYA, including speakers from Glasgow, Lothian (Edinburgh and surrounding areas) and both Orkney and Shetland (Insular Scots), stratified by gender (49 female, 46 male), age (49 aged 18-25, 46 aged 65+) and area (19 Insular, 28 Lothian, 48 Glasgow). Twice as much speech is taken from each speaker, with 180 seconds being considered for each speaker. In this larger corpus, I consider how voice quality varies according to social and linguistic factors, measured using f0-based coding of creak and H1*–H2*, H2*–H4*, H4*-2kHz* and CPP.

5.1 Previous work on Scottish voice quality

An early impressionistic remark on Scottish voice quality comes from Sweet (1902: 73):

> Narrowing of the upper glottis gives a wheezy character to the voice, sometimes approaching to strangulation. This effect is familiarly known as the 'pig's whistle.' It may be heard from Scotchmen, and combined with high key gives the pronunciation of Saxon Germans its peculiarly harsh character.

In Sweet's (1902: 12) explanation, 'wheeze' is one of the 'throat sounds', which he describes as 'that hoarse, wheezy sound known as 'wheezing' or 'stage whisper', which occurs when 'we strongly exaggerate an ordinary whisper'. Sweet posits that its production involves 'contraction of the upper glottis'. He transcribes it with a combination of the symbol for tenseness combined with the symbol for whisper. Laver (1980: 131) suggests that Sweet's noting of 'the pig's whistle' in 'Scotchmen' may refer to 'ventricular voice' or 'whispery ventricular voice', though I would suggest that Sweet's description could also be taken as simply whispery voice.

Other impressionistic remarks on Scottish voice quality include Catford (1977: 103) who noted that some degree of 'anteriorness' (rough equivalent to tense voice) is typical of North East Scots dialects, while very lax, full-glottal (roughly equivalent to modal voice) voice is typical of Central Scotland. Wells (1982: 82) remarks that lowland Scottish speakers typically use tense voice.

Three studies discussed in Section 3.2 used VPA to quantify variation in voice quality in Scottish accents. An additional study by Beck (1988) used VPA to describe voice quality in speakers with and without Down Syndrome, and included mostly speakers with Scottish accents resident in Edinburgh, but due to its focus on organic rather than social variation, did not restrict participants on the basis of accent. Here, I summarise the findings of these studies with a view to developing predictions of how voice quality might pattern in the data selected from SCOSYA.

Esling (1978b) looked at variation in voice quality by age and socio-economic class in male speakers in Edinburgh. Esling (1978b) sampled 32 adult men aged 22-76 and 18 boys aged 8-9 from Pilton and Morningside, divided into three groups of socio-economic status. The sample is also described with two additional participants (33 adults and 19 boys) by Esling (1978a), though the results appear to include only 50 speakers. He found that judgements of creaky voice predominated in the highest socio-economic status group, Group I, and judgements of harsh voice predominated in the lowest socio-economic status group, Group III. Each phonation type was rated on a scalar degree of 0-3, where 0 represented 'Not applicable' and 3 represented 'Extreme'. Esling (1978a) notes the incidence as a sum of all the scalar degrees for each group rather than presenting mean scalar degrees. Esling (1978a) notes the incidence of creaky voice at 19 in Group I, 12 in Group II, and 7 in Group III. As Group I contained 10 speakers and Groups II and III as contained 11 speakers each, this allows us to infer that Group I speakers were rated on average 1.9 out of 3 for creaky voice, Group II an average of 1.1, and Group III an average of 0.6. Esling (1978a) gives the incidence of harsh voice as 2 for Group I (inferred mean degree = 0.2), 15 for Group II (inferred mean degree = 1.4) and 21 for Group 3 (inferred mean degree = 1.9). Some degree of whispery voice was present across groups, at an incidence of 11 for Group 1 (inferred mean degree = 1.1), 8 for Group II (inferred mean degree = 0.7) and 9 for Group III (inferred mean

degree = 0.8). A second VPA analysis of the same speakers (Esling 1978b: 146) found an increased incidence of whispery voice, particularly in Group II and Group III, so that whispery voice was slightly more common in the lower socio-economic groups. Modal voice was reported to be slightly more common in Group I and II.

Beck (1988) looked at voice quality in 50 speakers without voice problems aged 18-40 resident in Edinburgh. These speakers mostly had south-east Scottish accents, but data was collected as a point of comparison for speakers with Down Syndrome, and so no particular accent was investigated. However, this study is still included here as it was the first study to consider gender variation using VPA, including 25 male and 25 female speakers, and includes mostly speakers with Scottish accents. Beck (1988) found that male speakers were more creaky, at an average scalar degree of 2.44 compared to 1.40 in female speakers. Whispery voice quality was prevalent in the sample, with all speakers being rated at scalar degree 2 or higher. Some speakers showed use of tense voice, with mean ratings for laryngeal tension at 0.92. Harsh voice was also identified in some speakers, at a mean scalar degree of 0.34.

Stuart-Smith (1999a) used VPA to consider variation in voice quality in Glasgow. She established VPA profiles for 32 speakers in both read speech, in the form of word lists, and conversational speech, in the form of 45 minute conversations between same-sex pairs. Speakers were stratified by age (young 13-14 years, adults 40-60 years), gender (male and female) and social class (working class from Maryhill and middle class from Bearsden), with 4 speakers in each category. She identified settings as distinctive of a group where they were present in three out of four speakers in that group, and the group mean of its scalar degree was statistically significant. She found that tense, whispery voice characterised Glaswegian laryngeal voice quality across speech styles and speaker groups. In terms of gender, she found more creaky voice in male speakers than female speakers and more whispery voice in female speakers than male speakers. Gender variation interacted with speech style, with increased whispery voice in female speakers being more apparent in read speech. Whispery voice was a marker of working class conversational speech overall, but working class female speakers were more whispery than their male counterparts.

Beck & Schaeffler (2015) looked at voice quality in 76 adolescent speakers aged 12-18 across Aberdeen, Inverness and Dumfries. They also found a predominance of whispery voice, ranging from a minimum mean scalar degree of 1.81 in male speakers from Aberdeen, to a maximum mean scalar degree of 2.36 in male speakers from Dumfries. There was no gender difference found in terms of whispery voice, but male speakers were found to use more creaky voice and harsh voice. Male speakers were found to use mean scalar degrees of 1.69, 1.43, and 1.09 for creaky voice in Aberdeen, Inverness and Dumfries respectively, compared to mean scalar degrees of 1, 0.56, and 0.96 for female speakers. Similarly, for harsh voice, male speakers used mean scalar degrees

of 0.68, 0.5, and 0.73 in Aberdeen, Inverness and Dumfries respectively, compared to mean scalar degrees of 0.38 0.13 0.21 for female speakers. Confirming remarks made by Catford (1977: 103), they identified tense voice as a feature of voice quality in Aberdeen: Larynx tension was significantly higher in Aberdeen speakers compared to Inverness. Mean scalar degrees for laryngeal tension were 0.82 for female speakers and 0.77 for male speakers in Aberdeen, compared to a range of 0.08 to 0.55 for other gender/area groups. Outside of this, they find no significant regional differences in terms of laryngeal voice quality.

While voice quality has never been studied systematically in Orkney and Shetland, Insular Scots dialects spoken in Orkney and Shetland today show differences from mainland varieties in terms of vocabulary, syntax (Smith & Durham 2011), phonetic realisation (Sundkvist 2011) and suprasegmental features (van Leyden & van Heuven 2006). Inhabitants of Orkney and Shetland spoke a variety of Norse, called Norn, until Scots was introduced from the mainland in the 15th century and gradually became the dominant language of the islands by the 18th century (Barnes 1984). Voice quality of Insular Scots has been noted to be distinct from mainland varieties of Scots, though most impressionistic remarks of this kind have concentrated on supralaryngeal settings. Catford (1957: 72) remarked that a kind of palatalisation exists in these varieties, describing it as having been a historical feature of certain consonants, but which he notes in the dipthongisation of vowels. Barnes (1991: 438) notes not being able to detect this in modern Shetland speech, but remarks that in a personal communication Gunnel Melchers suggested that 'statements on palatalisation are very much based on the dentalised articulatory setting'. Johnston (1997: 448) suggests that Insular Scots voice quality 'seems to be atypical of Scots'. Comparing remarks by Melchers' and Barnes (1984) on the voice quality of Shetland to Esling's (1978b) description of voice quality in Edinburgh leads Johnston (1997: 448) to suggest that Orkney and Shetland likely lack some of the distinctive features of voice quality in Edinburgh, specifically raised larynx and pharyngealisation. van Leyden & van Heuven (2006) note the presence of creak in Insular Scots speakers in their study on intonation, and exclude one participant due to excessive use of creak.

5.2 Predictions

5.2.1 Regional variation

I expected to find some broad differences in laryngeal voice quality between different Scots dialect areas of Scotland. Previous research seems to suggest that whispery voice is prevalent across Scottish accents, being found at least to some degree in Edinburgh (Esling 1978b); south-east Scottish accents (Beck 1988), Dumfries, Aberdeen and In-

verness (Beck & Schaeffler 2015); and Glasgow (Stuart-Smith 1999a). Because of this, I expect to find a prevalence of whispery voice in this study.

I expect to find some instance of harsh voice, particularly in Edinburgh, as this was identified by Esling (1978b). Following Stuart-Smith (1999b), I expect that harsh voice will not be particularly prevalent in Glasgow.

I expect to find tense voice in Glasgow (Stuart-Smith 1999b). Esling (1978b) did not differentiate between laryngeal and supralaryngeal tension, so it is difficult to formulate predictions in terms of tense voice for Edinburgh, but in south-east Scottish accents more generally, Beck (1988) found tense voice. Prevalence of tense voice may vary by region, as found by Beck & Schaeffler (2015). The only study to directly compare different Scottish accents is Beck & Schaeffler (2015), who find regional differences in laryngeal tension but not in any other aspect of phonatory quality.

Voice quality has not been studied systematically in some of these varieties (e.g. Orkney, Western Isles), so it is difficult to predict whether they will pattern in line with mainland varieties or exhibit their own distinct phonation profiles. In places that have been studied previously, I expect voice quality is likely to show similar characteristic features similar to those found in previous research, with whispery voice occurring in Glasgow and harsh voice and whispery voice occurring in Edinburgh. However, in the case that the present study uncovers findings that are inconsistent with previous research, this may suggest that laryngeal voice quality has changed over time in these varieties. If findings are inconsistent, differences between older and younger speakers might reveal whether this is due to a change over time — though as explored below in Section 5.2.3, care must be taken if attributing findings to changes over time rather than ageing (Hejná & Jespersen 2022). Alternatively, in line with Beck & Schaeffler (2015), it is possible that there may not be great difference in phonation between different varieties.

5.2.2 Gender

I also expect to find some gender variation in laryngeal voice quality, but anticipate that it is likely that the exact manifestation of this may not be consistent across all three accents studied here. Although direct comparison between different studies is difficult, Beck (1988), Beck & Schaeffler (2015) and Stuart-Smith (1999b) found differences in which phonation types show variation according to gender. Beck (1988) and Beck & Schaeffler (2015) found no significant differences in terms of whispery voice, while Stuart-Smith (1999b) found female speakers were more whispery in Glasgow. Beck (1988) and Stuart-Smith (1999b) found no gender differences in terms of harsh voice, while Beck & Schaeffler (2015) found that male adolescents used more harsh voice

than female adolescents and Esling (1978b) found harsh voice was characteristic of working class male speakers in Edinburgh. I there for expect that male speakers may be more creaky and/or harsh than female speakers, and that female speakers may be more whispery than male speakers, but I do not expect this to necessarily be consistent across all areas.

5.2.3 Age

Previous age comparisons in studies on Scottish voice quality have focused on differences between adults and children, while the present study compared younger and older adults. The younger age group for this study is aged 18-25, so participants will be expected to have undergone the laryngeal changes that occur during puberty.

The older group is aged 65+. In this age group, certain physiological processes that affect the larynx may have an effect on fundamental frequency, phonation duration, loudness and voice quality, known as 'presbyphonia' (Kendall 2007). Laryngoscopy of ageing vocal folds shows 'bowing' of the vocal folds and increased prevalence of a posterior glottal gap, thought to be attributable at least in part to laryngeal muscle atrophy (Martins et al. 2015). de Araújo Pernambuco et al. (2015) reviewed four studies of the prevalence of voice disorders in ageing populations and identified a prevalence of between 4.8% to 29.1%. Hannaford et al. (2005) gathered data on self-reported voice problems by age in Scotland, and found that 'croakiness' and 'loss or weakness' of voice significantly increase at the age of 75 and above. In a review and meta-analysis of the effects of ageing on the voice, Rojas, Kefalianos & Vogel (2020) found that older individuals were rated higher for dysphonia, roughness, breathiness, strain and instability on clinical auditory-perceptual scales, and that this manifested in increased noise acoustically. However, increased prevalence of voice disorders in older speakers does not necessarily indicate a universal decrease in vocal function with age in all speakers: Sapienza & Dutka (1996) considered voice performance in terms of glottal airflow characteristics in 60 women aged 20-70 years old who did not have voice disorders, and found no decrease in vocal function with age.

Ageing effects also appear to interact with speaker sex, as well as language or ethnic background. Rojas, Kefalianos & Vogel (2020) suggest that increased noise among male speakers was more well-established than it was for female speakers, and a number of studies find that when female speakers are considered, the situation may be more complex. Similarly, Gittelson, Leemann & Tomaschek (2021) finds opposite results in terms of acoustic measures for male and female speakers, with HNR35 and H1*–H2* lowering with age for male speakers but increasing for female speakers. Furthermore, Lee et al. (2016) note that most previous work on ageing voices concentrates on European American speakers of English, and gives tentative evidence that potentially

language background or ethnicity may play a role, finding that elderly Korean women use a tenser voice quality than younger Korean women, as shown by decreased H1–H2 and H1-A1 with age. As Hejná & Jespersen (2021, 2022) emphasise, the biological ageing process and social aspects of age can interact in their effects on the voice, so care must be taken at the interpretation stage if any effect is to be attributed solely to biological or social age.

In the absence of previous research specifically considering age variation in Scottish adults and the effects of ageing on the voice in Scotland, specific predictions are difficult to formulate. However, I expect there will be differences between older and younger speakers, likely manifesting in noisier voice quality in older speakers; For example, increased prevalence and scalar degrees of harsh, breathy and whispery voice. It is also possible that I will find changes in Scottish voice quality in comparison to previous studies; In this case, I would expect older speakers to pattern in line with trends found in previous studies but any novel findings on Scottish voice quality to be present among younger speakers.

5.2.4 Linguistic factors

As noted in Section 3.1, variation in voice quality can occur as a factor of linguistic context. Here I discuss the relevant linguistic contexts that may affect voice quality in Scottish accents in more detail.

The first of these factors is glottal stops, which are common in Scottish accents. Stuart-Smith (1999a) found that glottal stops were used for /t/ in 76% of cases in conversational speech in a study with 32 speakers stratified by age, gender and social class in Glasgow. Similarly, Schleef (2013) found glottals used in 85% of cases in conversational speech in 21 adolescents from Edinburgh. Glottals have also been reported in both Orkney and Shetland (Sundkvist 2011, Schmitt 2015). Additionally, glottal stops can occur for /p/ and /k/ in Glasgow, with McCarthy & Stuart-Smith (2013) finding glottals in 10% of cases for /k/ in word-final position in casual speech. In cases where oral stops are produced for /p/, /t/ and /k/, pre-glottalisation can also favour creak preceding these stops in word-final position (Gordeeva & Scobbie 2013). I expect to find increased prevalence of creak in the context of glottal stops, particularly in the context of glottalised /t/.

Creak is prevalent across accents of English as a marker of phrase-final position. Henton & Bladon (1988) found that creak was more common towards the end of an utterance in RP and Modified Northern, and this has also been found in Scottish English and American English speakers (J. Pearce 2019, Garellek 2015, Epstein 2002, Garellek, Ritchart & Kuang 2016, Kreiman 1982, Abdelli-Beruh, Wolk & Slavin 2014,

Wolk, Abdelli-Beruh & Slavin 2012).

Creak can also occur in phrase-initial vowel onsets; Following Dilley, Shattuck-Hufnagel & Ostendorf (1996), I refer to this as glottalisation. Pierrehumbert & Talkin (1992) considered where word-initial glottalisation occurred adjacent to vowels in two American English speakers in an experiment where phrase position, prosodic accent and stress were varied. They found that both stress and phrase position were factors: Stressed syllables showed noticeable glottalisation regardless of whether or not they occurred at a phrase boundary, while reduced syllables rarely showed noticeable glottalisation unless they occurred at the beginning of a phrase, while phrase accent affected the degree of glottalisation observed. Dilley, Shattuck-Hufnagel & Ostendorf (1996) consider glottalisation in 3709 word-initial vowels produced by two male and three female American English speakers. They found that speakers glottalised word-initial vowels more often at the beginning of an intonational phrase. They found accent played a lesser role than stress and phrase position, being a significant factor for non-phrase-initial full vowels and phrase-initial reduced vowels, but not in cases where glottalisation was already very likely (phrase-initial full vowels) or very unlikely (non-initial reduced vowels). I therefore expect to find increased incidence of creak in phrase-initial vowel onsets, as well as other acoustic cues to glottalisation in these contexts.

Several forms of aspiration noise originating from surrounding segments may also affect voice quality. The first is pre-aspiration, which occurs in Scottish English as before word-final voiceless fricatives (Gordeeva & Scobbie 2010). Gordeeva & Scobbie (2010) looked at pre-aspiration in five male and five female Scottish English speakers and found systematic presence of glottal aperiodic energy preceding word-final voiceless fricatives, supported by acoustic evidence from H1*–H2* and Harmonics-to-Noise ratio among other measures. Turk, Nakai & Sugahara (2006) and Gordeeva & Scobbie (2010) note that pre-aspiration can cause problems with segmentation; In the acoustic analysis, stretches were not hand-corrected to exclude aspiration noise, which may mean that pre-aspiration is included in the voiced stretch and affects acoustic measurement and perceived quality.

The second kind of aspiration likely to affect voice quality is that from preceding voiceless obstruents, which may occur following aspirated voiceless plosives /p,t,k/, or a preceding glottal fricative. I expect to find decreased incidence of creak and increased evidence of acoustic cues to glottal laxness and aspiration in these contexts.

Not all linguistic factors that could potentially affect voice quality are considered in the present study. However, by including random factors for the words contained in a voiced stretch in statistical models, I hope to control for other potential contributing factors. Stress and prosodic accent (Pierrehumbert & Talkin 1992) are two factors that

may affect voice quality which are not controlled for here.

5.3 Overview of corpus study

The corpus study is reported in four chapters. Chapters 6, Chapter 7 and Chapter 8 all involve a smaller sample of 24 participants from SCOSYA, stratified by age, gender, and area, with 90 seconds of speech from each speaker that was carefully checked for background noise. Chapter 9 then applies these methods to a larger sample.

Chapter 6 presents an auditory-perceptual analysis of Scottish voice quality using PPA. This method involves applying auditory-perceptual descriptive labels to voiced stretches, which are stretches of voiced sounds, usually sonorants. The purpose of this analysis is to anchor subsequent analysis in the descriptive terminology use in previous on Scottish accents that uses VPA, provide some insight into the patterns that might arise when more speakers are considered in the acoustic analysis, and allow an investigation of how these auditory-perceptual labels relate to automatic f0-based coding of creak in Chapter 7 and to acoustic measures in non-creaky phonation in Chapter 8.

Chapter 7 compares auditory-perceptual coding of creak, conducted at the level of the voiced stretch, to automatic f0-based coding of creak, conducted at the level of individual glottal pulses. The chapter investigates the issues and potential of f0-based coding of creak, attempts to understand the places where f0-based coding diverges from auditorily coded creak, and considers the feasibility of applying this f0-based coding of creak in the corpus analysis as a coarse-grained method of separating creak from non-creaky voice.

Chapter 8 compares acoustic analysis of voice quality in non-creaky voice to auditory-perceptual labelling of voice quality using PPA. This chapter aims to investigate patterns of acoustic measurement that mark specific voice qualities identified in auditory-perceptual analysis and consider whether there is an acoustic basis upon which whispery voice, breathy voice, tense voice and tense whispery voice can be differentiated.

Chapter 9 then scales the automatic f0-based coding of creak and acoustic analysis of non-creaky voice up to a larger sample of speakers from the corpus. 95 speakers from Glasgow (n=48), Lothian (n=28) and Orkney & Shetland (n=19) were considered, including 48 female speakers and 46 male speakers, as well as 49 younger speakers and 46 older speakers. The amount of speech from each speaker was increased to 180 seconds per speaker. This analysis first investigates variation in f0-based creak according to linguistic and social factors. As the f0-based method is used to separate creaky from non-creaky speech, this then allows an investigation of acoustic variation

in voice quality in non-creaky speech.

Chapter 6

Auditory-perceptual analysis using PPA

6.1 Introduction

In this section, I present a Phonation Profile Analysis (PPA) of 24 participants from SCOSYA (Smith et al. 2019), stratified by age, gender, and area. This analysis forms part of the wider investigation of Part II into voice quality variation by gender and age across three Scots varieties (Glasgow, Lothian and Insular Scots). The PPA method was developed with a view to grounding the analysis that follows in the larger corpus in the descriptive terminology that characterises previous VPA research on Scottish accents.

In Section 6.2, I outline the PPA method in more detail. I direct readers interested in the background to this method to Section 4.1, where I also outline the theoretical principles that guide this method, which are inspired by VPA. To recap, this method draws heavily from VPA, but focuses exclusively on phonation. Focusing on phonation allows smaller stretches of speech to be considered, in turn allowing consideration of how shorter-term variation in phonation compounds to produce overall laryngeal voice quality. It therefore takes stretches of speech that are maximally susceptible to variation in phonation as its unit of analysis, termed 'voiced stretches' (often referred to simply as 'stretches'), to consider how voice quality varies within and between speakers.

This analysis had several goals. Firstly, the analysis presented here allows Chapter 7 and Chapter 8 to compare the auditory-perceptual analysis to automatic f0-based coding of creak and acoustic analysis of non-creaky voice using multiple measures. This allows these methods to be rolled out to more speech and more speakers in Section 9 to investigate this variation acoustically, in a way that is grounded in descriptive phonetic terminology for voice quality. Secondly, I aimed to observe some general patterns according to social and linguistic factors. Doing so allowed development of more specific predictions as to how voice quality might vary according to social factors in acoustic analysis in with more data and speakers presented in Chapter 9. It also demonstrates the potential uses of this method in future research. Lastly, I aimed to assess the feasibility of conducting acoustic analysis on spontaneous speech collected outside of a speech lab, as previous work using the same approaches has tended to use on high-quality lab recordings.

More general predictions about how I expect voice quality to pattern in the corpus have just been outlined in the previous chapter (Chapter 5), but here I present more specific expectations about what I may find in PPA. I expected to find whispery voice to be prevalent across Scottish accents. I expected to find some variation between different areas in line with the findings of previous research, such as harsh voice in Edinburgh (Esling 1978b) and tenseness in Glasgow (Stuart-Smith 1999a). I expected that prevalence and degree of whispery voice, creak and harsh voice may vary by gender. I also expected to find differences between older and younger speakers, with

increased prevalence and scalar degrees of harsh, breathy and whispery voice among older speakers.

Not all the linguistic factors that might affect voice quality are analysed here. This is because this analysis is exploratory and the objective of considering variation according to linguistic factors is to demonstrate the usage of this method, rather than present an exhaustive analysis of all the linguistic influences on voice quality. I restrict my focus here to two factors that have previous been found to affect voice quality: Phrase position and the presence of a glottal stop. I expect that occurring in phrase-final position or in the context of a glottal stop will favour the use of creak.

Table 6.1: SCOSYA corpus demographics

	Female	Male	Total
Older (65+)	165	122	287
Younger (18-25)	173	116	289
Total	338	238	576

6.2 Methods

6.2.1 Overview of the corpus

6.2.1.1 Introduction

The Scots Syntax Atlas (SCOSYA) (Smith et al. 2019) corpus consists of 275 hours of recorded and transcribed data from 281 sociolinguistic interviews collected in approximately 2016. The corpus includes male and female speakers from 13 regions across Scotland, who fall into younger (18-25) and older (65+) age groups. Female speakers are over-represented, but the large number of speakers in the corpus means that it is still well-suited to considering gender variation in voice quality among young and old speakers across Scotland. No information is given on transgender status, and no non-binary speakers are identified.

6.2.1.2 Locations

Smith et al. (2019) identified potential dialect sub-regions in Scotland by using the Office of National Statistics' 'Travel To Work Area' maps, then choosing target areas based on population density and an assessment of dialect diversity. The corpus therefore contains speakers from 13 regions with distinct strands of Scots, including Doric in the North East, varieties spoken across the Central Belt, the Insular Scots of Orkney and Shetland, as well historically Gaelic-speaking regions in the Highlands and Islands. Te inclusion of areas with diverse linguistic histories makes th corpus well-suited to studying variation in voice quality across Scotland.

6.2.1.3 Participants

Smith et al. (2019) collected data through fieldworkers with an insider status each community. The fieldworkers and interviewed participants in pairs with a peer from the same community and age group. Demographics are given in Table 6.1. Pairs could be either same-gender or mixed-gender.

Participants were also recruited to meet a set of standard sociolinguistic criteria to

ensure that their speech was representative of the community. These criteria were:

- Born and brought up in the area;
- No significant time away from the area;
- Parents were also from the area;
- Had not gone on to higher education.

While the majority of the participants in the study met these criteria, interview context reveals this is not the case for every participant. Some participants talk about attending higher education, or mention being a few years outside of the target age group. However, complete demographic criteria were not available, so participants were presumed to still be close to the target demographics.

6.2.1.4 The sociolinguistic interviews

Participants took part in a sociolinguistic interview (Labov 1984) which aims to elicit spontaneous speech data and record information about participants' lives and community. However, in practice many conversations deviated from the format of a typical sociolinguistic interview; Interviewers and participants often seemed to know each other, and so often actual conversations involved catching up about recent events in each other's lives and gossip about mutual friends. Topics discussed include jobs, school, local life, films and sport. The interviews were conducted in the participants' homes to elicit naturalistic data in a comfortable setting. Each interview lasts approximately one hour.

6.2.2 Selection of smaller sub-corpus

As the SCOSYA corpus consists of 275 hours of data, it would be too time-consuming to force-align and conduct an auditory-perceptual analysis of voice quality on the entire dataset.

A sub-corpus was therefore selected where a limited amount of speech was taken from speakers from a restricted number of regions.

6.2.2.1 Selection of locations

The corpus was restricted to three of the areas in the full corpus. Initially, the objective had been to select distinct dialect areas which varied in terms of their linguistic

histories, as described by Johnston (1997) and Wells (1982), taking recordings from the Western Isles, Orkney & Shetland (Insular Scots), the North East, Lothian (in and near Edinburgh), and Glasgow. However, due to time constraints, recordings were only taken from Lothian, Glasgow, and Shetland.

6.2.2.2 Selection of speakers

Two speakers from each age and gender category were selected from each location, resulting in a total of 24 speakers, as shown in Table 6.2.

	Age		Gender		
Area	Younger (18-25)	Older (65+)	Male	Female	Total
Glasgow	2	2	2	2	8
Lothian	2	2	2	2	8
Shetland	2	2	2	2	8
Total	6	6	6	6	24

Table 6.2: Sampling of participants for subcorpus by gender, age and area.

Where possible, speakers were prioritised from pairs that were recorded together. In addition, I selected recordings with minimal background noise. Age is given at the time of recording, around 2016.

6.2.2.3 Selection of speech to analyse

Initially, I had hoped to constrain topic effects across recordings, but this was not possible due to variation between interviews: for example, some participants spent most of their interview discussing football, while other participants never mentioned this. Following Gregersen, Nielsen & Thøgersen (2009)'s discussion of selecting comparable data from different interviews, I then considered selecting portions of interviews on the basis of comparable interactional structure. However, this also proved impractical, as different participants had very different conversational styles: some interviews involved participants talking in long monologues, while others involved more speaker turns and interruptions. Because of the variation in topics and interactional structure between interview, I automated the selection using Script 1, which looks for the longest turn uttered by each speaker, then extracts 90 seconds of speech that includes the longest turn and speech preceding or following it until 90 seconds of speech is reached. The script then converts the sound file to mono, extracts the relevant speaker tier, and downsamples the sound file to 16 kHz, thus preparing the sound file and TextGrid for forced alignment. Where the automated procedure selected a portion of the interview with a large amount of background noise, I hand-selected a different 90 second extract from elsewhere in the interview.

6.2.3 Forced alignment procedure

In order to select voiced stretches (See Section 6.2.4), the data was force aligned using the Montreal Forced Aligner (McAuliffe et al. 2017b) v. 1.1. with the English acoustic model and English pronunciation dictionary. The acoustic model was trained on the LibriSpeech corpus, which contains speech from audiobooks spoken in a variety of accents, but predominantly American English (Panayotov et al. 2015). Forced alignment issues were anticipated due to phonological and lexical features of Scots. Scottish accents contain a single vowel for TRAP-BATH, while General American contains a FATHER-BOTHER merger, meaning that BATH and THOUGHT vowels are both represented with AA, despite not being produced with the same vowel in Scottish accents. However, as MacKenzie & Turton (2019) show, forced alignment can still be used reliably on varieties of British English, including Scots, which differ considerably from standard English, as long as alignment is hand-checked. An aligner based on General American instead of than Standard Southern British English had the advantage of allowing rhoticity to be represented in the phonemic transcription.

Alignment was therefore hand-checked, and places where the alignment was majorly displaced were either roughly corrected by hand or discarded from the data. More precise hand-correction was conducted at the beginnings and ends of voiced stretches, specified below.

Furthermore, many Scots words, neologisms, uncommon words, and product/brand names were not present in the English pronunciation dictionary. Many of these words were therefore added to the dictionary by hand, but some were omitted due to time constraints.

6.2.4 Selection of voiced stretches

Voiced stretches were selected as the unit of analysis for investigating voice quality. To be included in the analysis, stretches of speech met the criteria of including only:

- Phonologically voiced sounds, even where these sounds were realised without voicing (e.g. whisper)
- Sounds that were realised as voiced, regardless of whether they were phonologically voiced, where 'voiced' was defined as the presence of glottal pulses (e.g. creaky realisations of /t/ were includes as 'voiced')
- No segmental frication noise (e.g. frication noise in frication, burst or aspiration of plosive)

Sounds that met these criteria were considered to be maximally susceptible to effects of phonation (Laver 1980), with the exclusion of segments containing segmental frication noise being related to possibility of this noise being confused for aperiodic noise arising from the larynx. In practice, this meant that 'voiced stretches' tended to be composed of sonorants, i.e. vowels, approximants and nasals. Segments that were not phonologically sonorants were included if they were realised as such, such as plosives and fricatives being included where they were realised as approximants. Phonological sonorants were also excluded if they did not meet the above critera, for example, rhotics realised as taps or fricatives, or /w/ was realised as a voiceless labio-velar fricative.

Script 2 was used to extract stretches of consecutive vowels, approximants and nasals from the force aligned data. The boundaries of the voiced stretches were then hand-corrected. I took a conservative approach to excluding segmental noise from surrounding segments favouring excluding part of a vowel over including part of an adjacent fricative. This can be seen in Figure 6.3, where creaky cycles continue into a following /z/, but the boundary for the end of the voiced stretch is placed at the onset of visible friction noise from the frictive.

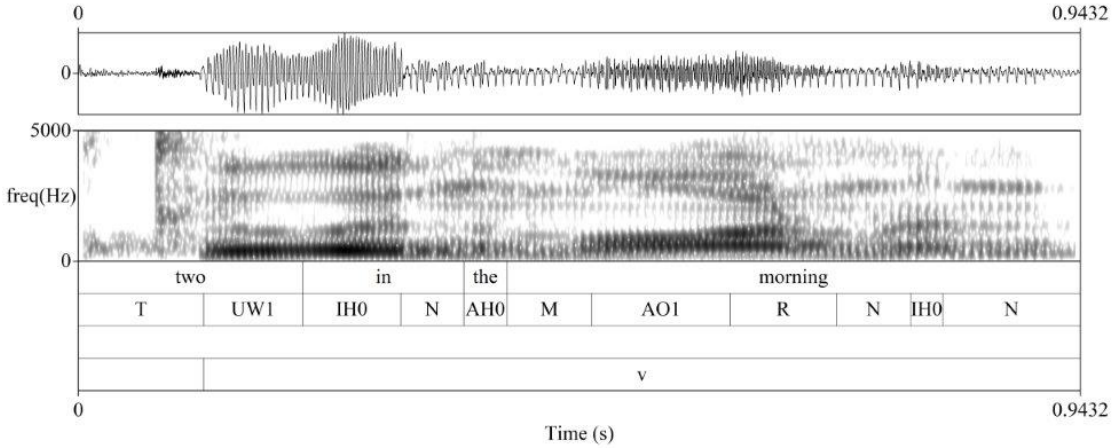


Figure 6.1: An extract from Graham YM Shetland, with a voiced stretch labelled 'v' marked on Tier 4.

Script 2 also adds tiers for each phonation type, preparing the TextGrid to be used to record the phonation types present in each voiced stretch. An example of an extract with a voiced stretch outlined is shown in Figure 6.1.

Stretches were excluded if there was speaker overlap, background noise, or were less than 100 ms, as it was considered too difficult to judge the speaker's voice quality in these cases.

Glottal stops required careful attention, as these were often realised as portions of creaky voice. If a glottal stop with full closure could be identified from the spectrogram and waveform, this was segmented in the phones tier along with other segments, as shown in Figure 6.2. However, in cases where a glottal stop was possible, but realised

as a portion of creak or tenseness, this was coded on a point tier below the phones tier, as shown on 6.3.

Portions of creaky voicing originating from the realisation of /t/ as a glottal stop were included within the boundary, and the presence of a glottal realisation of /t/ was coded. A boundary was then placed immediately after creaky voicing for the /t/ ceased, even if this was then followed by more voicing.

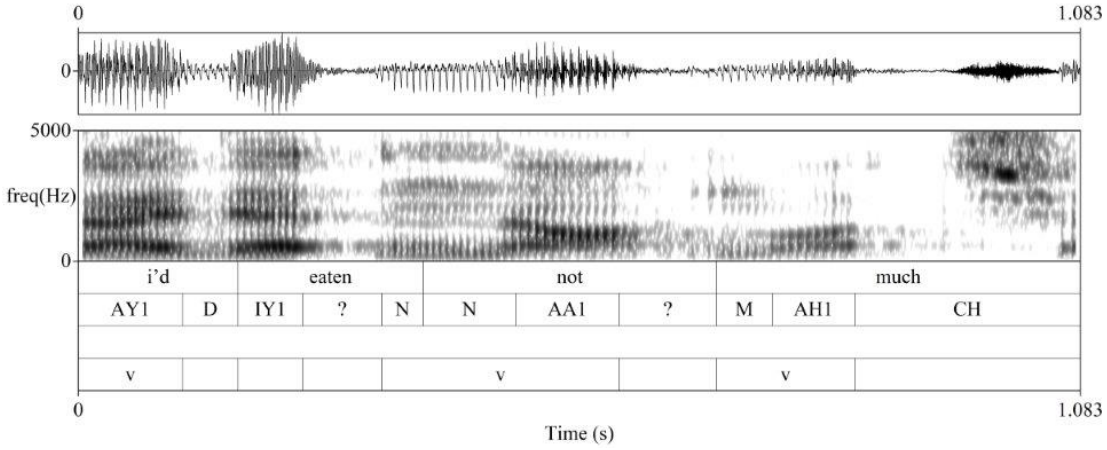


Figure 6.2: An extract from Graham YM Shetland, with glottal stops with full closure labelled '?' marked in an interval on Tier 2.

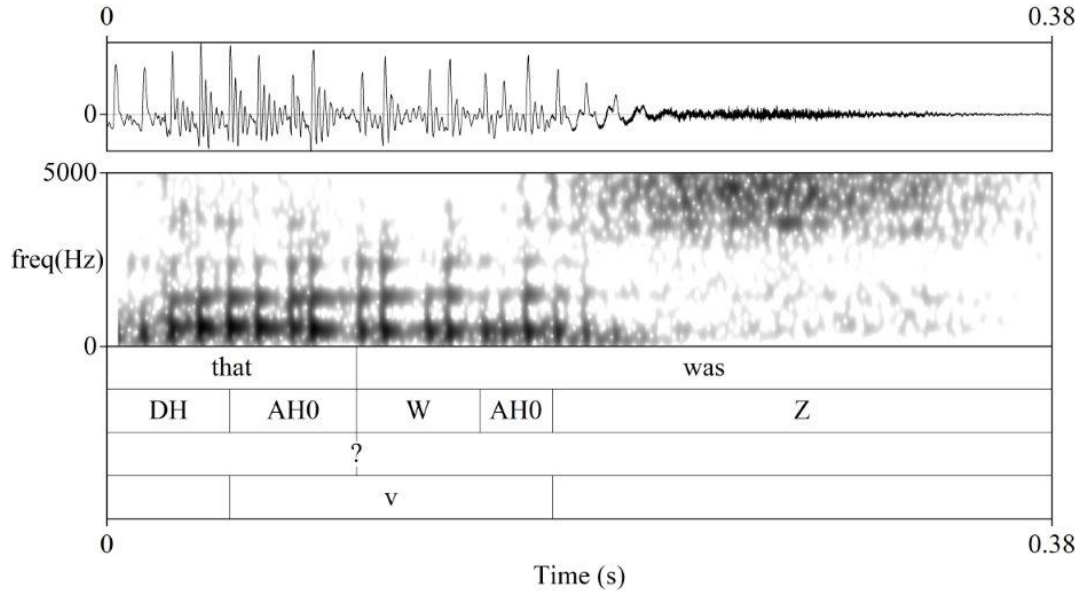


Figure 6.3: An extract from Scott YM Glasgow, with glottal stops realised as creak throughout the voiced stretch, and marked as '?' as a point on Tier 3.

Stretches were also separated into multiple shorter stretches if voice quality shifted dramatically during the stretch, as shown in Figure 6.4

However, notable phonation shifts that occurred towards the end or beginning of a chunk were excluded. For example, it was common for there to be just one or two cycles of creak at the beginning or end of a chunk, or a small amount of highly whispery

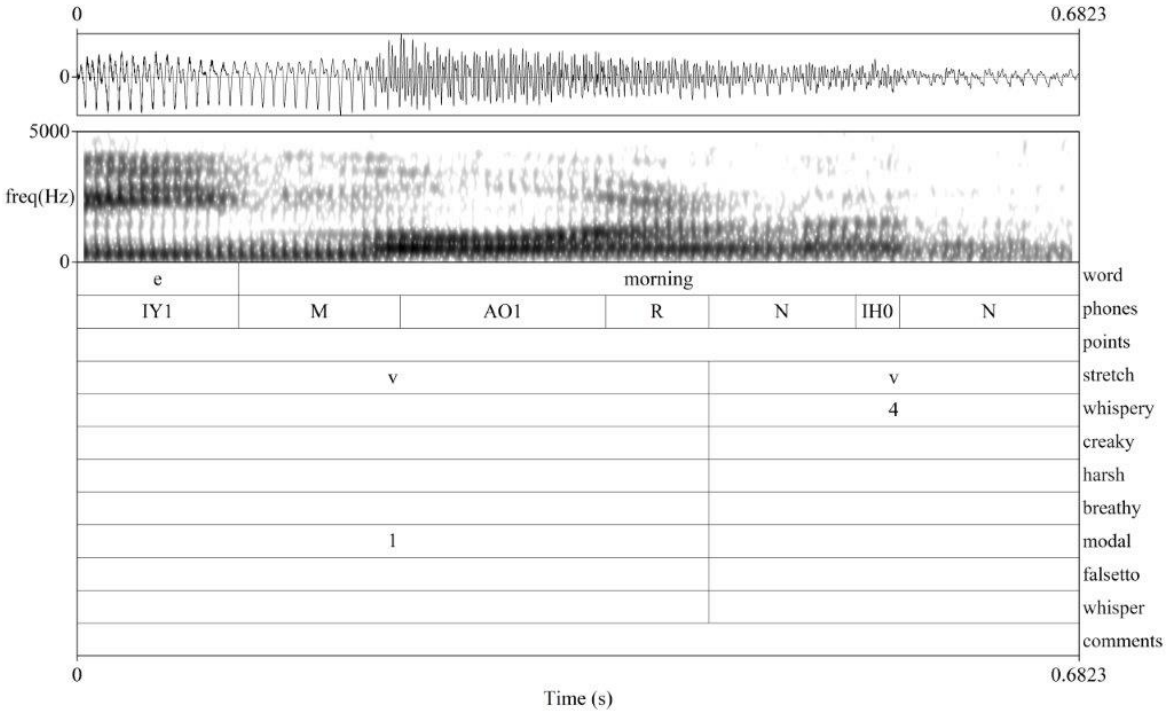


Figure 6.4: An extract from Graham YM Shetland, where an original voiced stretch is separated into two smaller chunks because voice quality changes from modal voice, coded as '1' on the modal tier, to whispery voice, coded as '4' on the whispery tier.

phonation before a voiceless fricative, and these were excluded unless they lasted more than 100 ms, making them eligible to be counted as a separate voiced stretch.

6.2.5 Auditory-perceptual analysis of phonation with Phonation Profile Analysis

As introduced in Section 4.1, auditory perceptual analysis of phonation type in voice stretches was conducted using PPA, a VPA-inspired method which allows consideration of how momentary variation between different phonation types and variation in auditory degree compounds to produce a speaker's overall laryngeal voice quality.

In this scheme, modal voice, whisper, harsh voice and falsetto were coded as binary, while whispery voice, creaky voice, and breathy voice were rated on scalar degrees of 0-5.

6.2.5.1 Development

Developing the system involved data analysis sessions with Jane Stuart-Smith and Felix Schaeffler, who were both trained in VPA and had used it in research. Sessions either involved myself coding test data ahead of the session, and the supervisor then

listening to a stretch and providing feedback, or both of us coding test data and then comparing our assessments and places where we agreed and disagreed. Over the course of the sessions, the number of scalar degrees for breathy, whispery, and creaky voice was refined, with 3-point, 5-point and 10-point systems tested before deciding on a 5-point scale. Different possibilities for compatibility of different phonation types were also considered, as was the possibility of separating 'creak' and 'creaky voice'. The procedure was further refined during analysis, as outlined below.

6.2.5.2 Modal voice

Modal voice is operationalised here as a 'neutral' reference setting and treated as binary. Following Laver (1980: 109-111) and Esling et al. (2019: 44-46), modal voice was coded where the following characteristics were apparent:

- Periodic vibration
- No audible friction resulting from incomplete glottal closure
- No audible friction resulting from epilaryngeal constriction

Additionally, to ensure that no instances of falsetto or creak were included as modal voice, I also paid careful attention to cases where fundamental frequency was notably above or below the usual range for a speaker, which could indicate a shift into creak or falsetto.

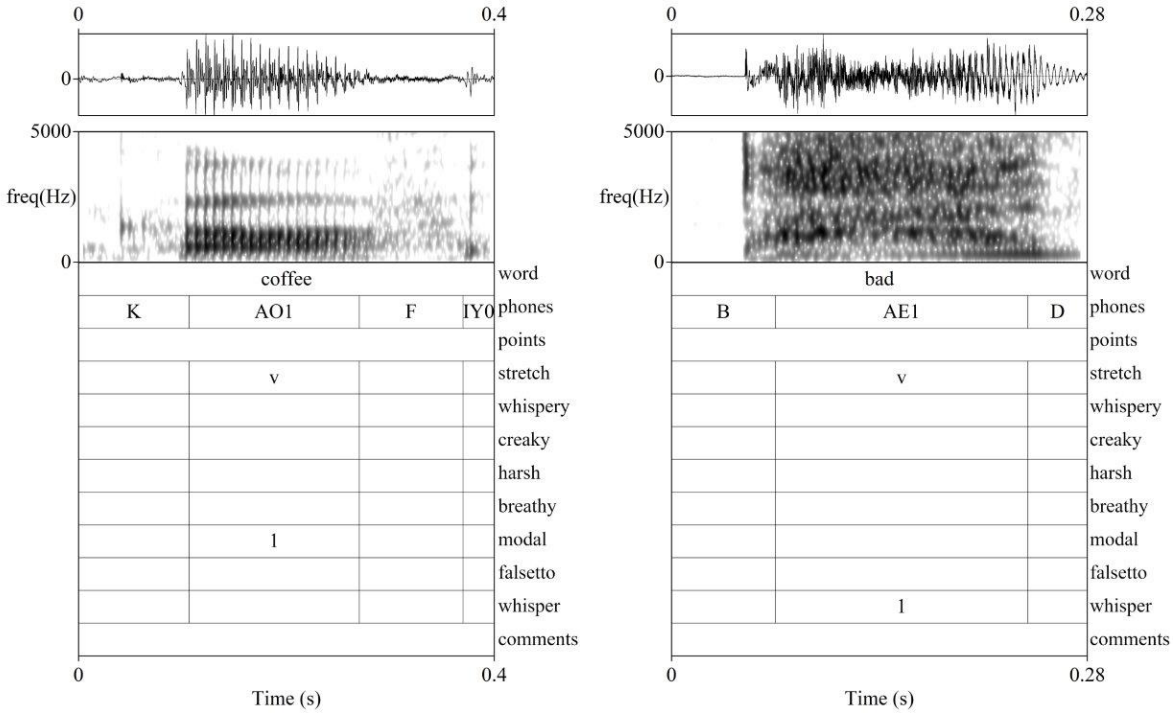
An example of modal voice is shown in Figure 6.5a. Notable characteristics identifiable in the spectrogram and waveform include periodic vibration at an f0 within this speakers usual range, clear formant structure, and little visible aperiodic noise in the spectrogram or waveform.

6.2.5.3 Whisper

Whisper was treated as binary and identified using the following criteria:

- A predominance of voicelessness during the stretch
- Usually accompanied by turbulent noise that dominates in the spectrogram or waveform

Diverging from Esling et al. (2019: 53-55), I did not make a distinction between the states of 'breath' and 'whisper'. This was due to voiceless states occurring infrequently



(a) Modal voice coded in the speech of Graham YM Shetland.

(b) Whisper coded in the speech of Jess YF Glasgow.

Figure 6.5: Examples of whisper and modal voice

and coding being conducted primarily auditorily, making any separation between the two laryngeal states difficult. Because of this, the criteria of turbulent noise is not essential - though airflow did appear to be turbulent in the majority of cases of whisper identified.

An example of whisper is shown in in Figure 6.5b. Notable characteristics of whisper visible here are aperiodicity in the waveform, visible noise in the spectrogram, unclear formant structure, and lack of voicing. Voicing does begin towards the end of the stretch, but auditory the effect of whisper dominates this stretch.

6.2.5.4 Falsetto

Falsetto was treated as binary. Auditorily, falsetto is characterised by high pitch and a quality that is 'almost flute-like' in nature (Zemlin 1964: 155). As Laver (1980: 199) and Esling et al. (2019: 61-62) note, falsetto is often accompanied by a degree of breathy or whispery friction noise. These characteristics can be seen in the example presented in Figure 6.6, where f0 reaches a maximum of 491 Hz and the whispery component was strong enough that the stretch was also coded as 4 for whispery voice.

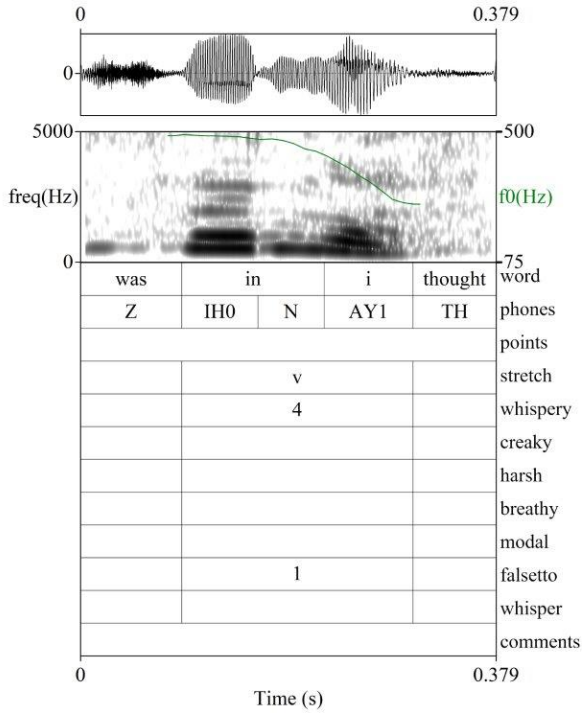


Figure 6.6: Falsetto coded in the speech of Alice YF Glasgow.

6.2.5.5 Harsh voice

Initially, harsh voice was coded on scalar degrees, but during the course of analysis, harsh voice was changed from a scalar quality to a binary one on the basis of few harsh tokens appearing in the test data. All tokens that were coded anywhere on the scale were then recoded as 1 for the presence of harsh voice. I coded any form of harsh voice outlined in Section 2.2.8 as harsh.

The following characteristics were used to judge the presence of harsh voice:

- Aperiodicity in fundamental frequency
- Additional aperiodic noise
- The categories of whispery voice and creaky voice are not sufficient by themselves to explain the quality - the stretch has an additional rough, raspy, or metallic sound not present in whispery voice or creak

6.2.5.6 Creaky voice

Because of the variability in manifestations of creaky voice, five scalar degrees were identified for creaky voice. The different scalar degrees correspond to different levels of auditorily creaky quality, as well as different types of creak identified in the literature.

Drawing on descriptions of creak by Laver (1980), Esling et al. (2019) and Keating, Garellek & Kreiman (2015), the scalar degrees were defined as followed:

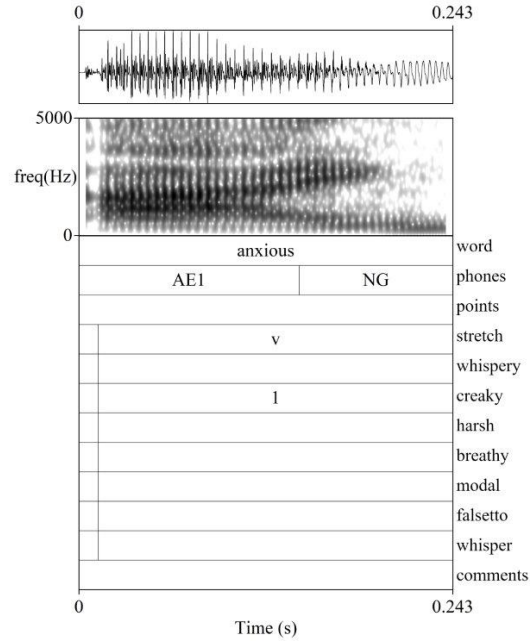
1. Tense voice: An auditory effect of glottal constriction, but lacking any other features of creak. Shown in Figure 6.7a.
2. 'Vocal fry' or periodic creak: f0 is notably lower than non-creaky speech of this speaker, but periodic. There is also a 'creaky' quality to the sound, and damped glottal pulses are visible in the waveform. Shown in Figure 6.7b.
3. Multiple pulsing: Two fundamental frequencies, visible in the waveform as a larger peak followed by a smaller peak. Shown in Figure 6.8a.
4. Prototypical or aperiodic creaky voice: Glottal constriction, with irregular f0. F0 either low or not identifiable due to irregularity. Auditory roughness and irregularity of f0 may mean that pitch is not auditorily low. Shown in Figure 6.8b.
5. Prototypical or aperiodic creaky voice: Glottal constriction, with irregular and low f0. F0 either low or not identifiable due to irregularity. But somehow 'more creaky' sounding than level 4, and has an auditorily low pitch. Shown in Figure 6.9.

During the course of coding, however, I became increasingly aware that differentiating these types of creak with such granularity was difficult to operationalise. Initially, I intended to treat aperiodic and prototypical creaky voice as two discrete categories, but quickly became aware that there was no particularly systematic criteria for me to differentiate them by in this data, as both were characterised by aperiodicity, and so I changed degree 4 and degree 5 to be differentiated based on a more auditory criteria. Furthermore, many instances of actual creak in this data included examples of multiple types of creaky voice within a single stretch, but not with any clear cut-off point.

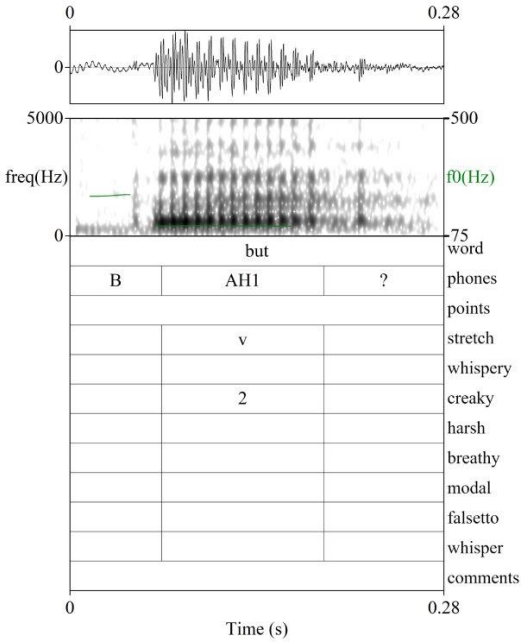
In the results section, I sometimes specifically refer to creaky voice scalar degrees 2-5 as 'creak', to mean any kind of creaky voice that contains alteration to fundamental frequency. This is used to differentiate it from tense voice.

6.2.5.7 Breathy voice

Previous VPA descriptions find Scottish accents to be whispery (Stuart-Smith 1999b, Beck & Schaeffler 2015). I therefore maintained a distinction between breathy and whispery voice. Following Esling et al. (2019: 56), breathy voice was taken to have a less turbulent airflow quality than whispery voice, as well as less auditorily constricted quality to its voicing.

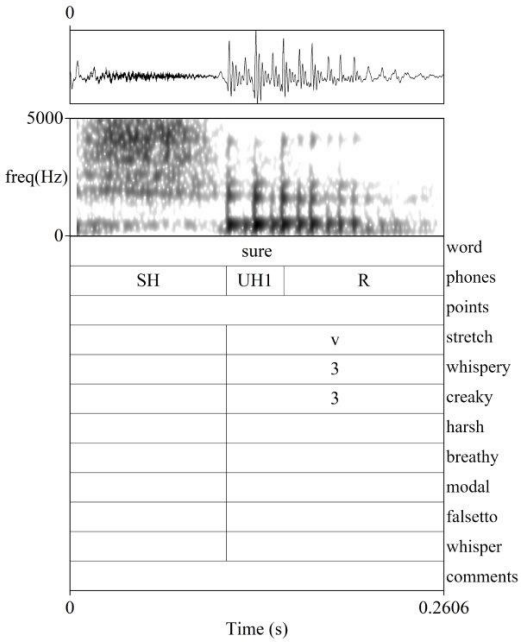


(a) Tense voice in the speech of Alice YF Glasgow. A single cycle of creak can be seen immediately preceding the stretch. There is little noise visible in the spectrogram and auditorily this sample resembles modal voice, but with an auditorily 'tense' quality. Coded as 1 in the creaky tier.

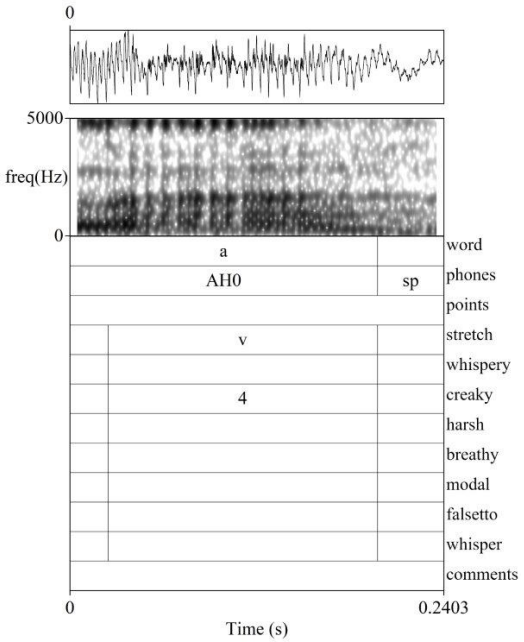


(b) Periodic creak in the speech of Kellie YF Shetland. F0 here is stable between 110-112 Hz, much lower than this female speaker's non-creaky f0 range. Coded as 2 on the creaky tier.

Figure 6.7: Examples of scalar degree 1 and 2 of creaky voice.



(a) Multiple pulsing in the speech of Scott YM Glasgow. A larger peak followed by a smaller peak can be seen in the waveform the first few cycles, though this does not persist throughout the stretch. Coded as 3 on the creaky tier.



(b) Aperiodic creak in the speech of Kellie YF Shetland. Though this stretch contains cycles of low f0 creak, aperiodicity and higher f0 cycles mean that this stretch does not have an auditorily low pitch. Coded as 4 on the creaky tier.

Figure 6.8: Examples of scalar degrees 3 and 4 of creaky voice.

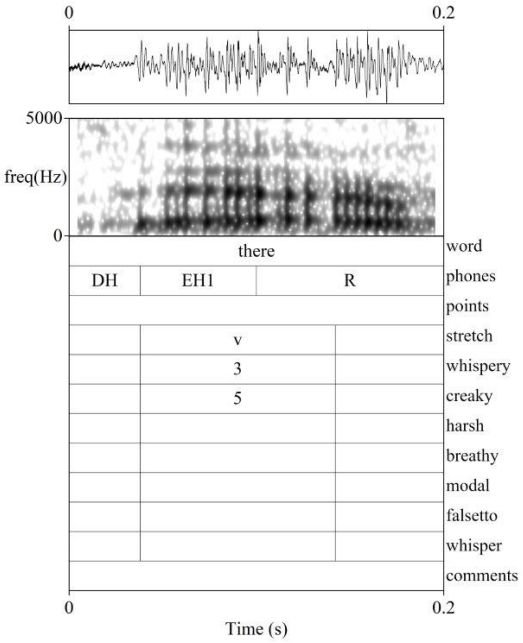
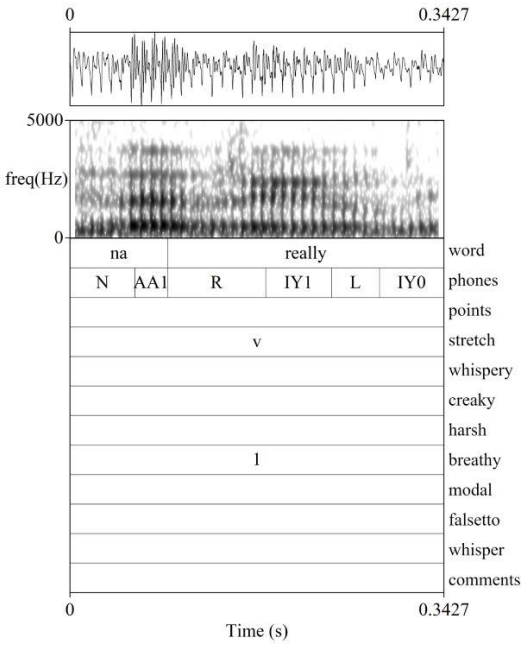


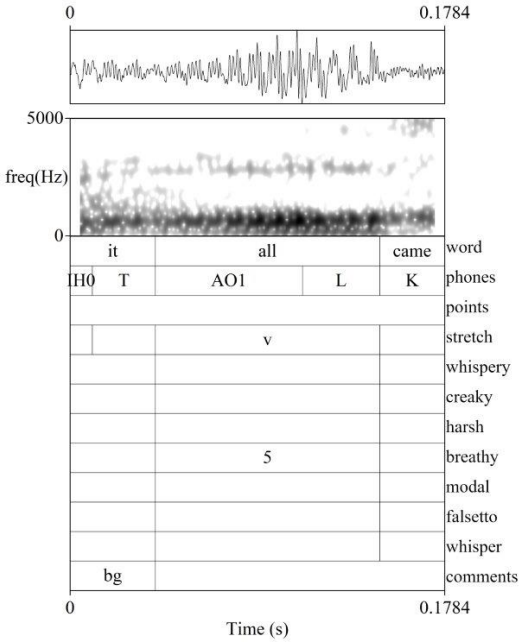
Figure 6.9: Prototypical creak in the speech of Kellie YF Shetland. Despite aperiodicity and containing both higher and lower f0 cycles, the auditory effect is a very low pitch. Coded as 5 on the creaky tier.

I followed this distinction, and identified breathy voice in cases where I heard a lax quality in the vocal folds, accompanied by a degree of noise that fit with the noise profile of examples of breathy voice produced by John Laver and examples from my own data that I had confirmed to be breathy with my supervisors during development of the protocol.

A scalar degree of 1 reflected what is often called 'slack voice' or 'lax voice', corresponding to an unconstricted vocal fold vibration with little or no audible friction. A scalar degree of 5 reflects a higher degree of audible friction, with an auditory quality of the vocal folds 'flapping in the breeze' (Catford 1977: 99).



(a) Breathy voice scalar degree 1 in the speech of Graham YM Shetland. There is a near modal auditory quality with auditory laxness



(b) Breathy voice scalar degree 5 in the speech of Gary OM Lothian. Formant bandwidth are wide and there is additional aperiodic noise, but not to the same extent as scalar degree 5 of whispery voice. At higher frequencies, there is very little spectral energy.

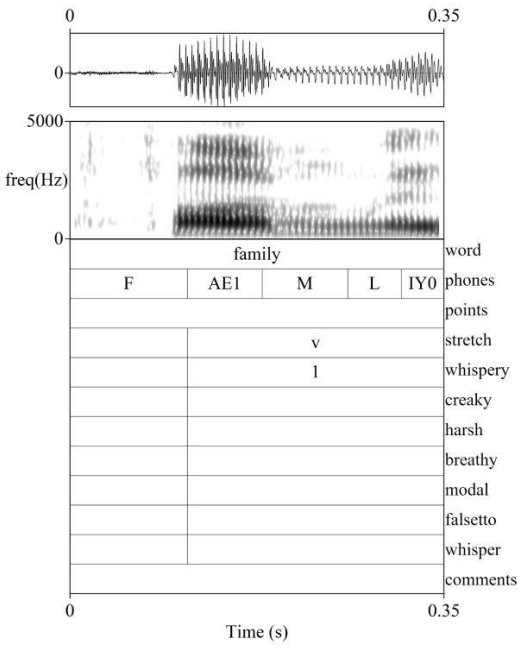
Figure 6.10: Examples of scalar degrees 1 and 5 of breathy voice

6.2.5.8 Whispery voice

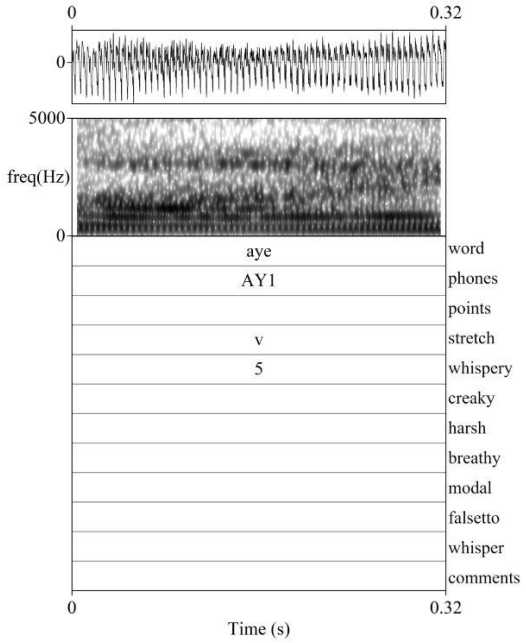
Following Esling et al. (2019: 58), I coded for whispery voice where I identified the following characteristics:

1. Audible glottal friction, usually greater than the amount present in breathy voice coded at an equivalent degree

2. No auditorily lax quality



(a) Whispery voice scalar degree 1 in the speech of Alice YF Glasgow. Near modal with some additional noise.



(b) Whispery voice scalar degree 5 in the speech of Alice YF Glasgow. Characterised by strong turbulent noise and aperiodicity, and unlike in breathy voice this noise is present in higher frequencies.

Figure 6.11: Examples of scalar degree 1 and 5 of whispery voice

Figure 6.11 shows examples of scalar degree 1 and 5 of whispery voice. Comparison between scalar degree 5 of whispery voice shown in Figure 6.11b and scalar degree 5 of breathy shown in Figure 6.10b demonstrates the different between these. The whispery example shows extreme aperiodic noise, while the breathy example shows comparatively moderate aperiodic noise, wide formant bandwidths, and decreased energy in higher parts of the spectrum.

6.2.5.9 Compound types

I followed Laver (1980) and Esling et al. (2019) when considering which phonation types could combine and which could not, as well as discussing examples of particular cases with my supervisors during development of the scheme.

Due to my treatment of modal voice as a baseline reference category, I treated it as unable to combine with any other phonation type. During development of the protocol, the 'modal voice' tier was instead marked 'voicing', and was checked when voiced phonation types were present (e.g. 'breathy voice' was coded with a mark in the 'voicing' tier and a mark in the 'breathy' tier), while true 'modal voice' was marked by only a mark in the 'voicing' tier. I see the distinction between these two options I considered as more of a practical decision than a theoretical one, as it meant that I

only had to make one mark rather than two to code other voiced phonation types.

Whisper was also taken as unable to combine with any other phonation type.

Whispery voice, creaky voice and harsh voice were taken as able to combine. A common combination was tense whispery voice, marked by a 1 in the creaky tier accompanying any degree of whispery voice. Due to the the fact that harsh voice is characterised by epilaryngeal constriction, many instances of harsh and creaky voice were realised as whispery.

I considered falsetto as able to combine with whisperiness, creakiness, harshness.

Contrary to Laver (1980), I considered breathy voice and creaky voice as able to combine. Previous research (Slifka 2006, Gobl & Chasaide 2000) have identified cases of lax creak, and I found cases of creak that appears to align with this, and some had additional aperiodic noise, leading me to code them as higher than just scalar degree 1 for breathy voice.

Breathy voice was taken to be incompatible with harsh voice or whispery voice.

6.2.6 Analysis

Due to the small number of participants in the smaller subcorpus sample, I did not attempt any statistical analysis of the data. Instead, analysis takes the form of exploring the distribution of each voice quality and scalar degree coded by social and linguistic factors.

The social factors considered here were age (older or younger), gender (female or male) and area (Glasgow, Lothian, Shetland).

Linguistic factors considered were phrase final position (final or not final) and glottal context (glottal or not glottal).

Glottal context was defined as a stretch containing or being immediately preceded or followed by a glottal stop. An example of a stretch containing a glottal stop is given in Figure 6.3. Phrase position was estimated based on whether the preceding and following segment on the phones tier was coded as a pause or blank.

The data was explored through graphical representations and contingency tables in `R v 4.1.1` (R Core Team 2020). Contingency tables were generated using the `janitor` package (Firke 2020) and exported using xtable (Dahl et al. 2019). Stacked percentage bar plots created using `ggplot2` (Wickham 2016) were used to visualise the proportion of each scalar degree used within a single voice quality. Euler diagrams were created

using `eulerr` (Larsson & Gustafsson 2018).

6.3 Results of PPA for 24 speaker sample

2170 voiced stretches were considered in the analysis. The distribution of phonation types across all voiced stretches is shown in Table 6.3 and visualised in the Euler diagram in Figure 6.12. Euler diagrams use area-proportional ellipses to represent each phonation type, with intersections between ellipses showing combination types. The placement of ellipses in relation to each other is arbitrary. Whispery creaky voice was the most common phonation type, followed by whispery voice, breathy voice, creaky voice and modal voice. Other phonation types were rare, with no other quality making up more than 2% of tokens.

Table 6.3: Distribution of phonation types across all voiced stretches

Phonation types	Freq	%
Breathy	370	17
Creaky	246	11
Creaky breathy	18	1
Creaky falsetto	1	<1
Creaky harsh	1	<1
Harsh	6	<1
Modal	126	6
Whisper	3	<1
Whispery	620	29
Whispery creaky	737	34
Whispery creaky falsetto	1	<1
Whispery creaky harsh	3	<1
Whispery falsetto	1	<1
Whispery harsh	37	2
Total	2170	100

Please note that audio examples for all figures presented in this chapter, and in later chapters, can be accessed by contacting the author of the thesis, Joe Pearce. It was not possible to make these available openly due to data originating from SCOSYA. It is also possible to request access to the spoken data for all of SCOSYA, `https://scotssyntaxatlas.ac.uk/about/accessing-the-spoken-corpus/`.

6.3.1 Overall distributions of each scalar degree

Whispery voice, breathy voice and creaky voice were rated on scalar degrees. Percentages and counts for each of these are given in Table 6.4.

6.3.1.1 Distributions of scalar degrees for simple phonation types

As shown in Figure 6.13, scalar degree 3 was the most common scalar degree for exclusively whispery voice (38%) and exclusively breathy voice (32%). Scalar degree

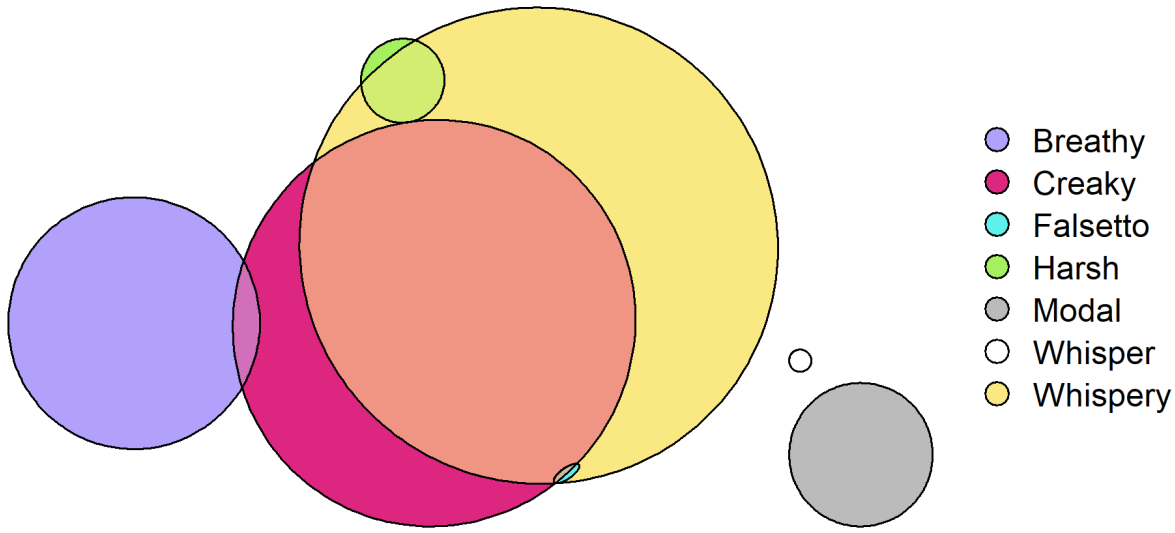


Figure 6.12: Euler diagram showing the distribution of voice quality categories in the data. Each ellipsis represents a discrete voice quality category; intersections between ellipses represent combination voice quality types. The area of ellipses and intersections is proportional to the number of data points in that category.

1, indicating lax voice was also very common for exclusively breathy voice (28%)

For exclusively creaky voice, the most frequently coded scalar degree was also 1, with 83% of exclusively creaky voice voiced stretches being coded as scalar degree 1 (i.e. tense voice).

Voice quality	**1** % (n)	**2** % (n)	**3** % (n)	**4** % (n)	**5** % (n)	Total % (n)
Whispery creaky						
Creaky	81 (595)	6 (44)	2 (18)	2 (18)	8 (62)	100 (737)
Whispery	26 (192)	41 (300)	21 (153)	9 (67)	3 (25)	100 (737)
Breathy	28 (104)	22 (82)	32 (120)	12 (45)	5 (19)	100 (370)
Whispery	15 (96)	30 (186)	38 (237)	12 (76)	4 (25)	100 (620)
Creaky	83 (205)	7 (18)	4 (10)	1 (3)	4 (10)	100 (246)

Table 6.4: Distribution of scalar degrees of whispery creaky voice, exclusively whispery voice and exclusively breathy voice across all voiced stretches. For each cell, the percentage is given first, followed by the count in brackets

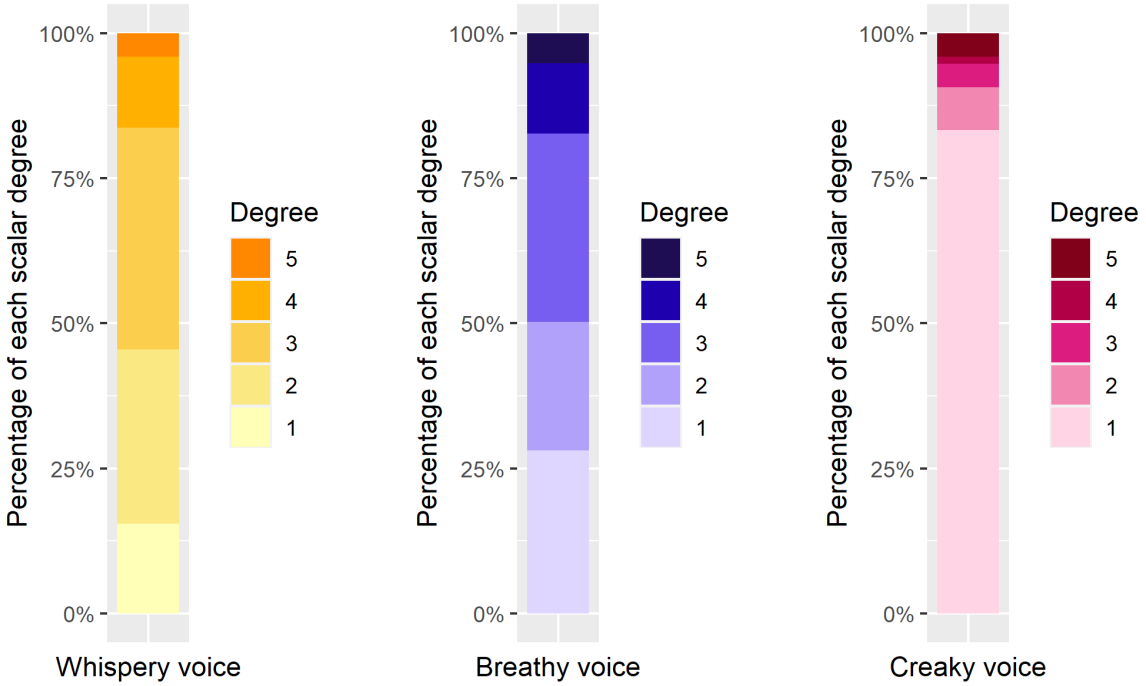


Figure 6.13: Stacked percentage bar plots showing the distribution of each scalar degree of whispery voice, breathy voice and creaky voice for voiced stretches coded as only one voice quality.

6.3.1.2 Distributions of scalar degrees for whispery creaky voice

As shown in Figure 6.14, where stretches were coded as whispery creaky voice, the most common scalar degrees were scalar degree 2 for whispery voice (41%) and scalar degree 1 for creaky voice (82%), indicating that this mostly consisted of tense whispery voice.

6.3.2 Social factors

In this section, I explore how PPA ratings of voice quality varied by social factors. Additional contingency tables can be found in Appendix A.1 where these are not shown.

6.3.2.1 Area

Speakers from different areas varied in terms of the proportion and scalar degree of phonation types they used.

Table 6.5 shows the distribution of each voice quality by area.

As shown through the orange intersection of the ellipses for whispery and creaky

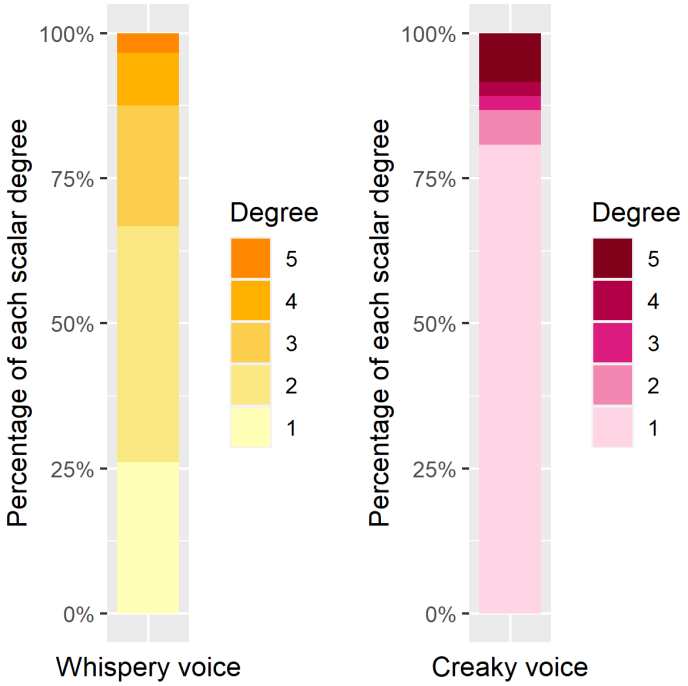


Figure 6.14: Stacked percentage bar plots showing the distribution of each scalar degree of whispery voice and creaky voice for voiced stretches coded as whispery creaky voice.

voice in Figure 6.15a, whispery creaky voice was the most frequent voice quality used speakers in Glasgow, appearing to characterise Glaswegian voice quality. The next most commonly used voice quality was whispery voice, followed by breathy voice, then creaky voice.

There was a wide distribution of different phonation types used by Lothian speakers, as shown by the relatively similar size of the breathy, whispery, creaky and modal ellipses in Figure 6.15b. Though breathy voice was most common, creaky voice, whispery creaky voice, whispery voice and modal voice were all prevalent among Lothian speakers. Though harsh voice was still uncommon, Lothian speakers used harsh voice and harsh combination types more often than speakers from other areas, shown in the green ellipsis.

Speakers in Shetland used whispery creaky voice most frequently, followed by whispery voice, creaky voice, breathy voice and modal voice.

We have seen that speakers in Glasgow used more whispery creaky voice than speakers from Lothian or Shetland. However, the scalar degrees favoured for each component of whispery creaky voice also vary by area. In Figure 6.16a, we can see that speakers in Lothian favour scalar degree 1 for both the whispery and creaky component, indicating that whispery creaky voice used by Lothian speakers tends to be closer to modal voice in comparison to other areas. Meanwhile, speakers in Shetland appear to favour higher scalar degrees of creaky voice in whispery creaky voiced stretches more than speakers from Glasgow or Lothian do. The distribution of each scalar degree by

Table 6.5: Distribution of all phonation types by area

Voice quality	Glasgow	Lothian	Shetland
Breathy	12 (92)	32 (214)	9 (64)
Creaky	4 (33)	21 (138)	10 (75)
Creaky Breathy	1 (10)	1 (7)	0 (1)
Creaky Falsetto	0 (1)	0 (0)	0 (0)
Creaky Harsh	0 (0)	0 (1)	0 (0)
Harsh	0 (1)	1 (5)	0 (0)
Modal	0 (2)	12 (78)	6 (46)
Whisper	0 (1)	0 (1)	0 (1)
Whispery	36 (276)	14 (92)	35 (252)
Whispery Creaky	45 (344)	17 (117)	38 (276)
Whispery Creaky Falsetto	0 (1)	0 (0)	0 (0)
Whispery Creaky Harsh	0 (0)	0 (3)	0 (0)
Whispery Falsetto	0 (1)	0 (0)	0 (0)
Whispery Harsh	1 (5)	3 (17)	2 (15)
Total	100 (767)	100 (673)	100 (730)

area is given in Table 6.6.

Table 6.6: Contingency table showing degree of whispery voice and creaky voice by social factor in voiced stretches coded as whispery creaky voice

		1	**2**	**3**	**4**	**5**	Total
Area	Component	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Glasgow	Creaky	84 (288)	4 (15)	2 (6)	3 (12)	7 (23)	100 (344)
	Whispery	24 (82)	46 (158)	19 (66)	8 (29)	3 (9)	100 (344)
Lothian	Creaky	89 (104)	2 (2)	3 (4)	1 (1)	5 (6)	100 (117)
	Whispery	37 (43)	43 (50)	15 (17)	4 (5)	2 (2)	100 (117)
Shetland	Creaky	74 (203)	10 (27)	3 (8)	2 (5)	12 (33)	100 (276)
	Whispery	24 (67)	33 (92)	25 (70)	12 (33)	5 (14)	100 (276)

We have already seen that speakers in Glasgow and Shetland used similar amounts of exclusively whispery voice, while speakers from Lothian used this quality less frequently. As shown in Figure 6.17a, speakers from different areas also varied in terms of which scalar degrees they favoured. When using exclusively whispery voice, Glasgow and Lothian speakers used very similar amounts of scalar degree 3, but differed in their use of other degrees. Glasgow speakers used more scalar degree 4 and 5 than Lothian speakers, while Lothian speakers used more of scalar degree 1 and 2. The distribution of scalar degrees used for whispery voice by area is given in Table 6.7

Table 6.7: Contingency table showing degree of whispery voice by area in exclusively whispery voiced stretches

	1	**2**	**3**	**4**	**5**	Total
Area	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Glasgow	7 (18)	25 (70)	47 (129)	16 (44)	5 (15)	100 (276)
Lothian	20 (18)	30 (28)	46 (42)	3 (3)	1 (1)	100 (92)
Shetland	24 (60)	35 (88)	26 (66)	12 (29)	4 (9)	100 (252)

As we have seen, breathy voice is more common in Lothian than in Glasgow. Figure

6.17b shows how the degree of breathy voice used varies by area. It shows how speakers in Lothian use more lax voice (breathy voice degree 1) than speakers in Glasgow or Shetland, with the lightest shade of purple occupying more space in the plot: Lothian speakers used this degree in 42% of breathy voiced stretches, compared to just 7% for Glasgow speakers. This continues a trend seen in whispery and whispery creaky voice of Lothian speakers using lower scalar degrees of each quality. The distribution of scalar degrees for breathy voice by area is given in Table 6.8.

Table 6.8: Contingency table showing degree of breathy voice by area in exclusively breathy voiced stretches by area

	1	**2**	**3**	**4**	**5**	Total
Area	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Glasgow	7 (6)	29 (27)	40 (37)	18 (17)	5 (5)	100 (92)
Lothian	42 (89)	21 (44)	29 (62)	7 (15)	2 (4)	100 (214)
Shetland	14 (9)	17 (11)	33 (21)	20 (13)	16 (10)	100 (64)

As we already saw, speakers in Lothian used the most exclusively creaky voice, followed by speakers in Shetland, and speakers from Glasgow, who used the least. As shown in Figure 6.17c, speakers from all areas used similar proportions of tense voice, with tense voice making up the highest proportion of creaky voice for all areas. However, speakers from different areas did appear to favour different scalar degrees of the higher degrees of creak. For example, Shetland speakers used more of scalar degree 2 than speakers from other areas, while Glasgow speakers used more of scalar degree 5.

6.3.2.1.1 Creaky voice by area including all combination types As creaky voice frequently occurred as a combination type, I also considered the distribution of creaky voice by area including all combination types. Table 6.9 shows the distribution of creaky voice by area, subsuming all stretches coded as creaky voice scalar degree 2-5 into 'creak' and scalar degree 1 into 'tense voice'. Presented in this format, we see that Shetland speakers use more creak than speakers from other areas, whereas Glasgow speakers use more tense voice than speakers from other areas. Lothian speakers, who used the highest rates of exclusively creaky voice, in fact use the lowest rates of creaky and tense voice across all combination types.

Table 6.9: Contingency table showing the distribution of creak (creaky voice scalar degree 2-5), tense voice (creaky voice scalar degree 1) and non-creaky voice by area, including all combination types.

	Creak	Tense	Not creaky	Total
Area	% (n)	% (n)	% (n)	% (n)
Glasgow	9 (69)	42 (320)	49 (378)	100 (767)
Lothian	7 (44)	33 (222)	60 (407)	100 (673)
Shetland	12 (86)	36 (266)	52 (378)	100 (730)

6.3.2.2 Gender

As summarised in Table 6.10 and shown in Figure 6.18, the overall proportions of each voice quality used varied by gender. As shown by the large orange intersection between the large whispery and creaky ellipses in Figure 6.18a, female speakers used more whispery creaky voice, whispery voice, and creaky voice than male speakers. On the other hand, the large purple circle in Figure 6.18b demonstrates that male speakers used more breathy voice than female speakers. Male speakers also used more modal voice and more harsh voice than female speakers, as shown by the increased size of the grey and green circles.

Table 6.10: Distribution of all phonation types by gender

Voice quality	F % (n)	M % (n)
Breathy	3 (37)	31 (333)
Creaky	16 (174)	7 (72)
Creaky Breathy	0 (0)	2 (18)
Creaky Falsetto	0 (1)	0 (0)
Creaky Harsh	0 (0)	0 (1)
Harsh	0 (2)	0 (4)
Modal	4 (49)	7 (77)
Whisper	0 (1)	0 (2)
Whispery	33 (356)	24 (264)
Whispery Creaky	43 (464)	25 (273)
Whispery Creaky Falsetto	0 (0)	0 (1)
Whispery Creaky Harsh	0 (0)	0 (3)
Whispery Falsetto	0 (1)	0 (0)
Whispery Harsh	1 (6)	3 (31)
Total	100 (1091)	100 (1079)

Though whispery creaky voice was more common among female speakers, the gender variation in scalar degrees used within whispery creaky voice was minimal, as summarised in Table 6.11. Across genders, speakers favoured a tense component (scalar degree 1) in whispery creaky voice, added to scalar degree 1-3 whispery voice.

Table 6.11: Contingency table showing degree of whispery voice and creaky voice by gender in voiced stretches coded as whispery creaky voice

gender	Component	**1** % (n)	**2** % (n)	**3** % (n)	**4** % (n)	**5** % (n)	Total % (n)
F	Creaky	83 (387)	6 (28)	2 (11)	2 (7)	7 (31)	100 (464)
	Whispery	29 (135)	43 (198)	20 (91)	7 (34)	1 (6)	100 (464)
M	Creaky	76 (208)	6 (16)	3 (7)	4 (11)	11 (31)	100 (273)
	Whispery	21 (57)	37 (102)	23 (62)	12 (33)	7 (19)	100 (273)

In exclusively whispery and creaky voice, however, female and male speakers also appeared to favour different scalar degrees of each voice quality.

While female speakers used more exclusively whispery voice overall, they tended to use lower scalar degrees for whispery voice than male speakers. Female speakers used more of scalar degrees 1-3 than male speakers, while male speakers used more of scalar degrees 4-5. This is shown by the prevalence of lighter yellows shades for female speakers in Figure 6.19a, and is summarised in Table 6.12.

Table 6.12: Contingency table showing degree of whispery voice by area in exclusively whispery voiced stretches by gender

	1	**2**	**3**	**4**	**5**	Total
Gender	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
F	19% (69)	32% (113)	39% (139)	7% (25)	3% (10)	100% (356)
M	10% (27)	28% (73)	37% (98)	19% (51)	6% (15)	100% (264)

As well as using more breathy voice overall, where male speakers used breathy voice they tended to use higher scalar degrees than female speakers. This is shown by the prevalence of lighter shades of purple for female speakers in Figure 6.19b, compared to the darker shades of purple among male speakers. Table 6.13 shows the distribution of each scalar degree of breathy voice by gender.

Table 6.13: Contingency table showing degree of breathy voice by area in exclusively breathy voiced stretches by gender

	1	**2**	**3**	**4**	**5**	Total
Gender	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
F	62 (23)	24 (9)	11 (4)	3 (1)	0 (0)	100 (37)
M	24 (81)	22 (73)	35 (116)	13 (44)	6 (19)	100 (333)

Though female speakers used more exclusively creaky voice overall, male and female speakers used similar proportions of scalar degree 1 creaky voice, as shown in Figure 6.19c. However, when using types 2-5 creaky voice, male speakers appeared to use almost exclusively scalar degree 2 creaky voice, while female speakers used a range of different scalar degrees 2-5. The distribution of different scalar degrees by gender is summarised in Table 6.14.

Table 6.14: Contingency table showing degree of creaky voice by gender in exclusively creaky voiced stretches

	1	**2**	**3**	**4**	**5**	Total
Gender	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
F	83 (145)	5 (8)	6 (10)	1 (2)	5 (9)	100 (174)
M	83 (60)	14 (10)	0 (0)	1 (1)	1 (1)	100 (72)

Table 6.15 shows the distribution of creak and tense voice by gender across all combination types. This shows that female speakers use far higher rates of tense phonation types than male speakers do, using tense qualities nearly twice as often as male speakers - a pattern that was not visible when only exclusively creaky stretches were considered. Rates of creak (scalar degree 2-5) are very similar across genders across all combination types.

Table 6.15: Contingency table showing the distribution of creak (creaky voice scalar degree 2-5), tense voice (creaky voice scalar degree 1) and non-creaky voice by age, including all combination types.

Gender	Creaky % (n)	Tense % (n)	Not creaky % (n)	Total % (n)
F	10 (106)	49 (533)	41 (452)	100 (1091)
M	9 (93)	25 (275)	66 (711)	100 (1079)

6.3.2.3 Age group

Compared to the influence of area and gender, age group appears to have a smaller influence on the voice quality of speakers, and the proportions of each voice quality appear quite similar between older and younger speakers. Figure 6.20 shows the distribution of phonation types across older and younger speakers. However, older speakers appear to use more harsh voice than younger speakers, shown in the green ellipses in Figure 6.20. In fact, younger speakers used harsh voice in less than 1% of voiced stretches. Older speakers also appear to use more exclusively whispery voice than younger speakers, as shown by the increased size of the yellow ellipsis. Furthermore, younger speakers appear to use more modal voice than older speakers, as shown by the increased size of the grey ellipsis.

Age group	Creaky n (%)	Tense n (%)	Not creaky n (%)	Total n (%)
O	7 (84)	38 (432)	54 (615)	100 (1131)
Y	11 (115)	36 (376)	53 (548)	100 (1039)

Table 6.16: Contingency table showing the distribution of creak (creaky voice scalar degree 2-5), tense voice (creaky voice scalar degree 1) and non-creaky voice by age, including all combination types.

There was little difference in the scalar degrees used by each age group for most qualities. However, as summarised in Table 6.16, younger speakers used more creak (scalar degree 2-5 of creaky voice) when this was considered across all combination types, despite younger and older speakers using similar amounts of tense voice.

6.3.3 Linguistic factors

In this section, I explore how the linguistic context affected PPA ratings of voice quality. For the sake of brevity, contingency tables are not given for linguistic factors here, but can instead be found in Appendix A.1.

6.3.3.1 Glottal context

As shown in Figure 6.21, glottal context appeared to influence the overall rate of creaky voice used by speakers.

There was more creaky voice and combination types of creaky voice and other qualities in glottal context, as shown by the larger pink circle in Figure 6.21a than in Figure 6.21b.

The scalar degree of creaky voice coded also tended to be higher in glottal contexts. This is shown in Figure 6.22 for the degree of creaky voice coded in creaky and whispery creaky stretches by glottal context.

This is also evident if we consider the proportion of non-creaky stretches, stretches coded scalar degree 1 (tense voice), stretches coded scalar degrees 2-5 of creaky voice ('creak') by glottal context across all combination types. As summarised in Table 6.17, there was considerably more creak in glottal contexts than non-glottal contexts, and slightly more tense voice.

Glottal context	Creaky n (%)	Tense n (%)	Not creaky n (%)	Total n (%)
Not glottal	6 (113)	37 (673)	57 (1045)	100 (1831)
Glottal	25 (86)	40 (135)	35 (118)	100 (339)

Table 6.17: Contingency table showing the distribution of creak (creaky voice scalar degree 2-5), tense voice (creaky voice scalar degree 1) and non-creaky voice by glottal context, including all combination types.

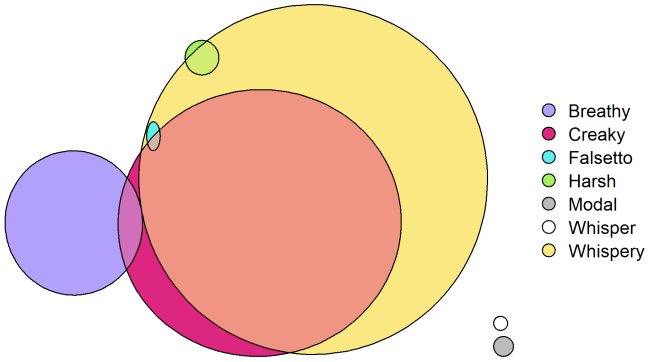
6.3.3.2 Phrase final position

As shown in Figure 6.23, the distribution of each voice quality was broadly similar between final and non-final phrase position. However, in final position, there were higher rates of exclusively creaky voice and modal voice than in non-final position. There were also lower rates of breathy and whispery voice in final position.

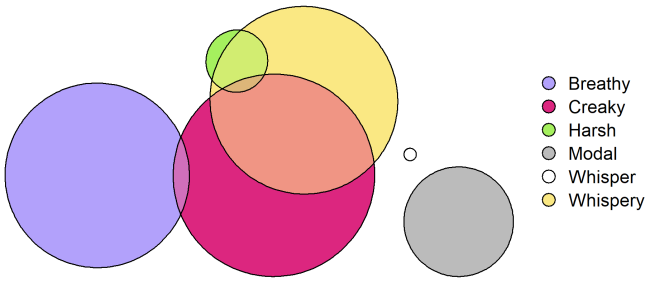
The scalar degrees of used for each voice quality were also very similar between final and non-final position. As summarised in Table 6.18, there was slightly more creak and tense voice in final position than non-final position, but this difference was small.

Phrase position	Creaky	Tense	Not creaky	Total
Not final	9 (144)	37 (615)	55 (911)	100 (1670)
Final	11 (55)	39 (193)	50 (252)	100 (500)

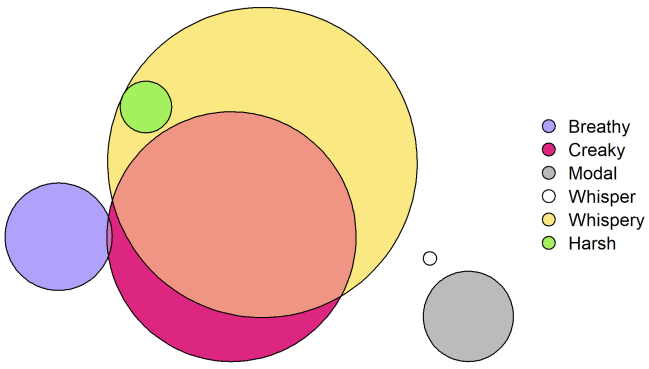
Table 6.18: Contingency table showing the distribution of creak (creaky voice scalar degree 2-5), tense voice (creaky voice scalar degree 1) and non-creaky voice by phrase position, including all combination types.



(a) Glasgow (n = 767)

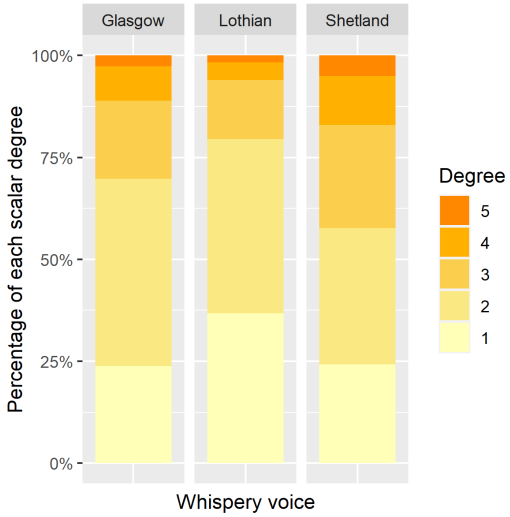


(b) Lothian (n = 673)

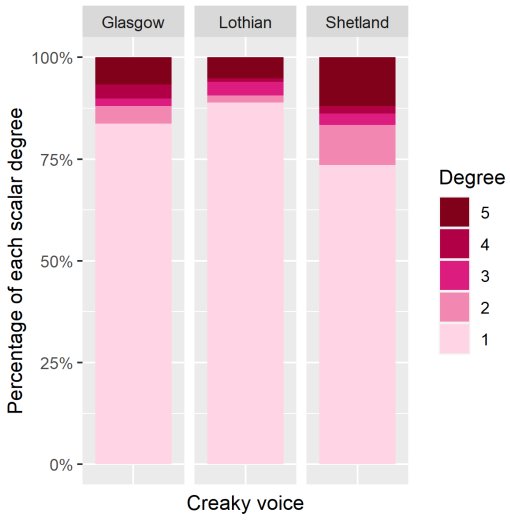


(c) Shetland

Figure 6.15: Euler plots showing the distribution of voice quality categories and how they combine in Glasgow, Lothian and Shetland

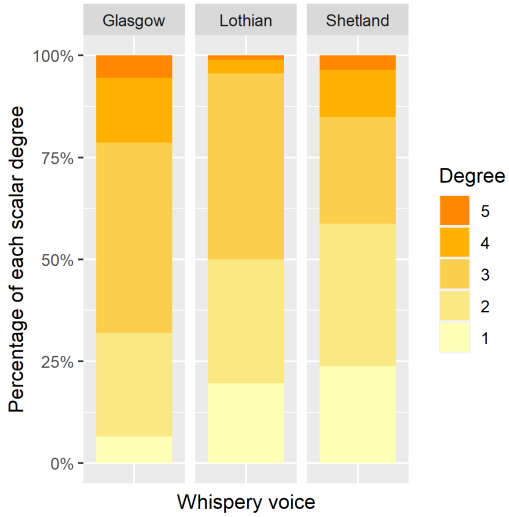


(a) Proportion of scalar degree of whispery voice for whispery creaky voiced stretches by area

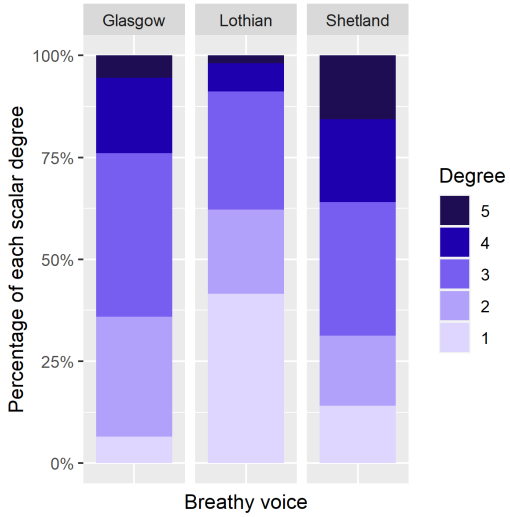


(b) Proportion of scalar degree of creaky voice for whispery creaky voiced stretches by area

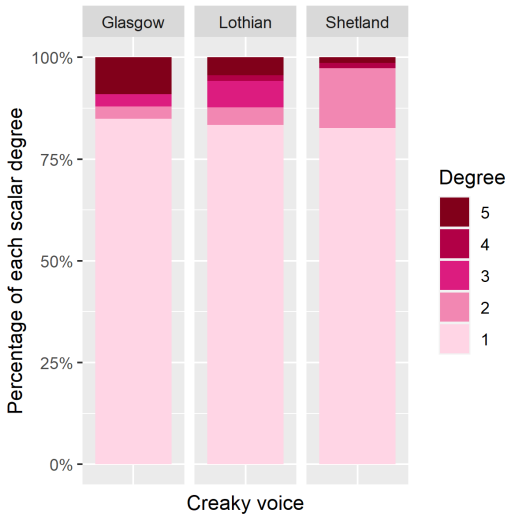
Figure 6.16: Stacked percentage bar plots showing the proportion of each scalar degree of whispery voice and creaky voice for voiced stretches coded as whispery creaky voice in Glasgow, Lothian and Shetland.



(a) Proportion of scalar degree of whispery voice for exclusively whispery voiced stretches by area

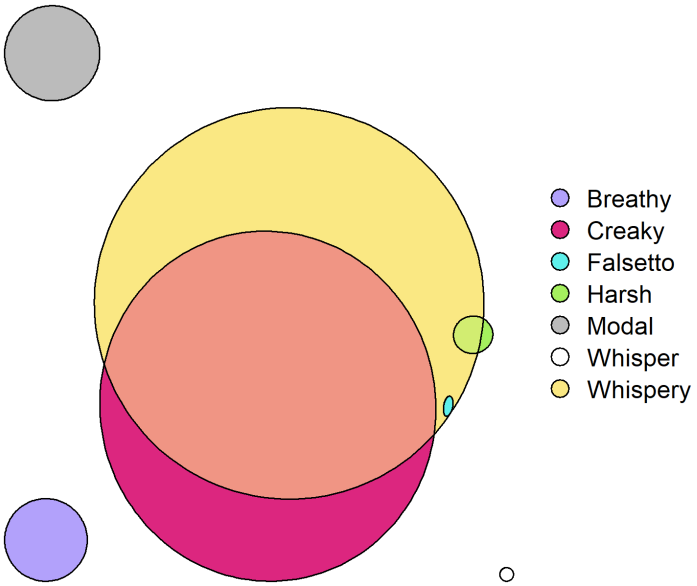


(b) Proportion of scalar degree of breathy voice for exclusively breathy voiced stretches by area

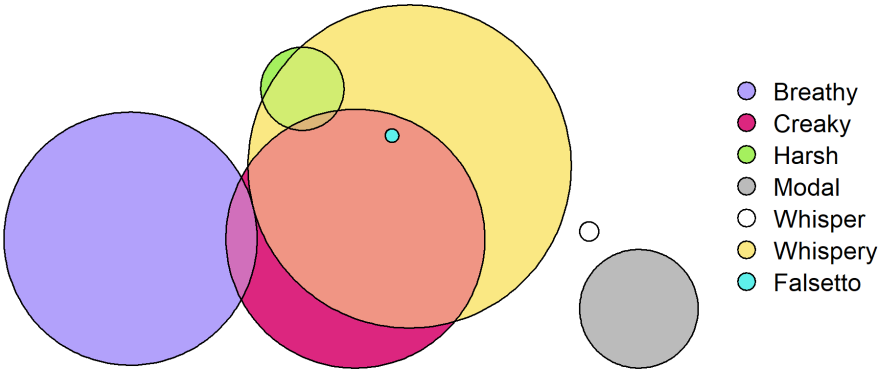


(c) Proportion of scalar degree of creaky voice for exclusively creaky voiced stretches by area

Figure 6.17: Stacked percentage bar plots showing the proportion of each scalar degree of whispery voice, breathy voice and creaky voice for voiced stretches coded as only as whispery voice, breathy voice and creaky voice in Glasgow, Lothian and Shetland.

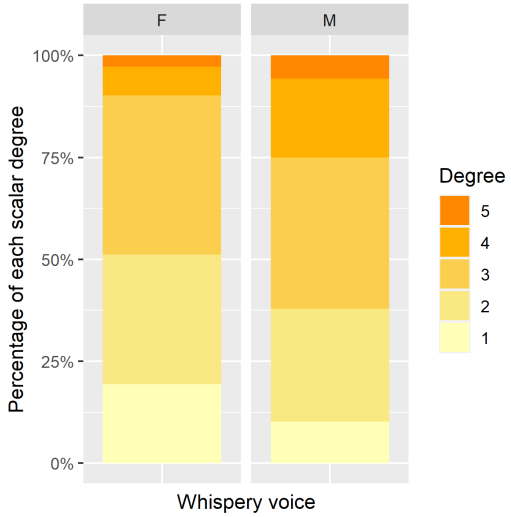


(a) Female speakers (n = 1091)

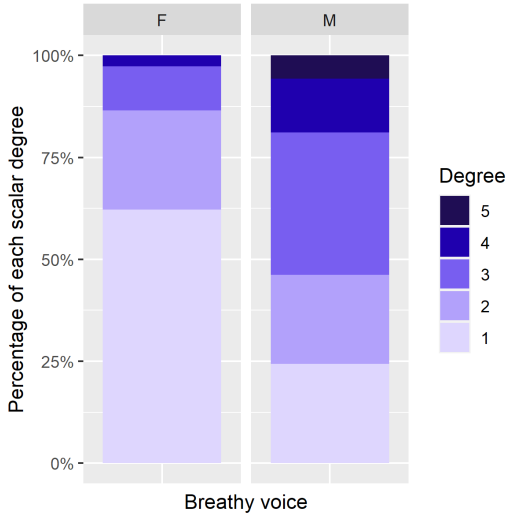


(b) Male speakers (n = 1079)

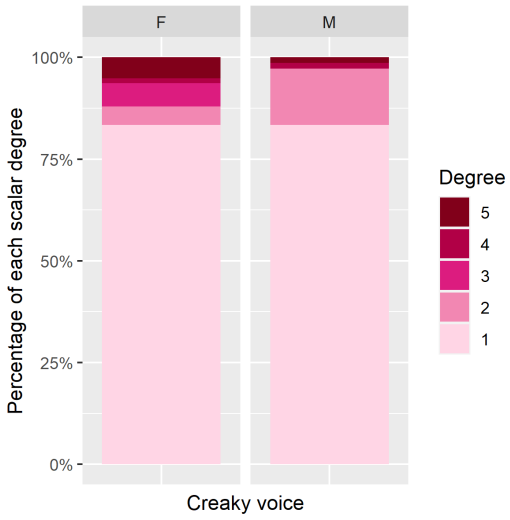
Figure 6.18: Euler plots showing the distribution of voice quality categories and how they combine in female and male speakers



(a) Proportion of scalar degree of whispery voice for exclusively whispery voiced stretches by gender

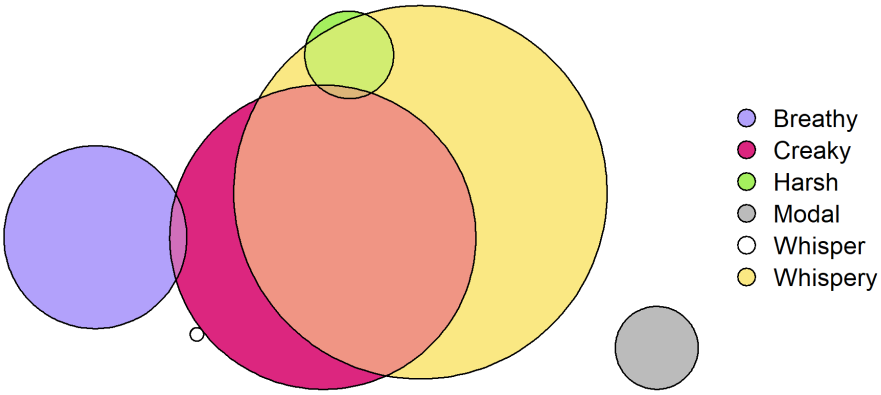


(b) Proportion of scalar degree of breathy voice for exclusively breathy voiced stretches by gender

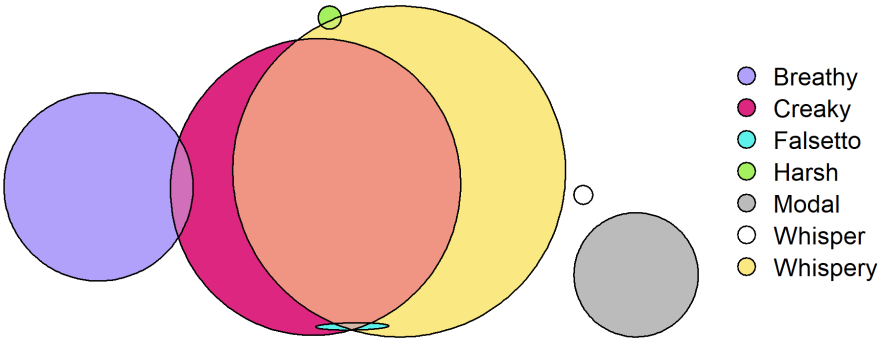


(c) Proportion of scalar degree of creaky voice for exclusively creaky voiced stretches by gender

Figure 6.19: Stacked percentage bar plots showing the proportion of each scalar degree of whispery voice and breathy voice and creaky voice for voiced stretches coded as only as whispery voice, breathy voice or creaky voice for female and male speakers.

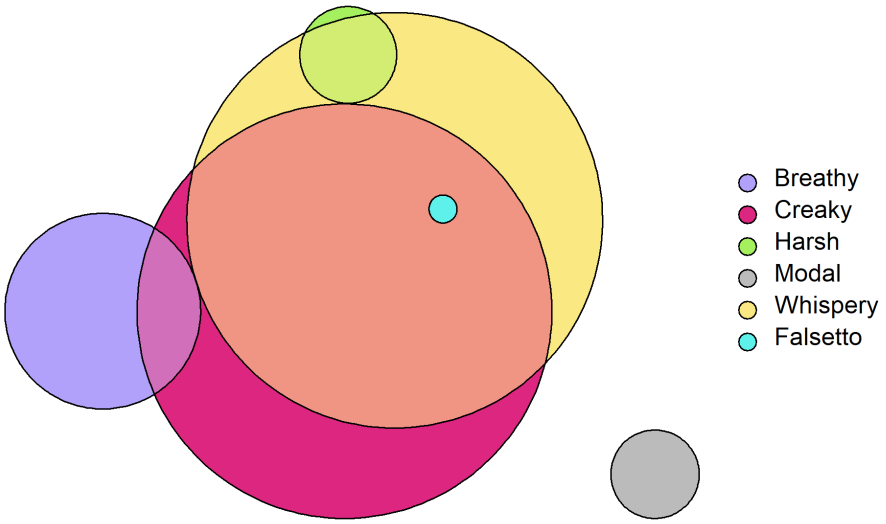


(a) Older speakers

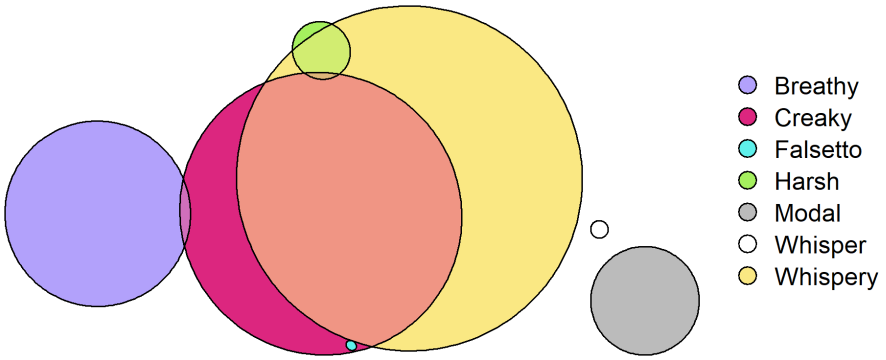


(b) Younger speakers

Figure 6.20: Euler plots showing the distribution of voice quality categories and how they combine in older and younger speakers

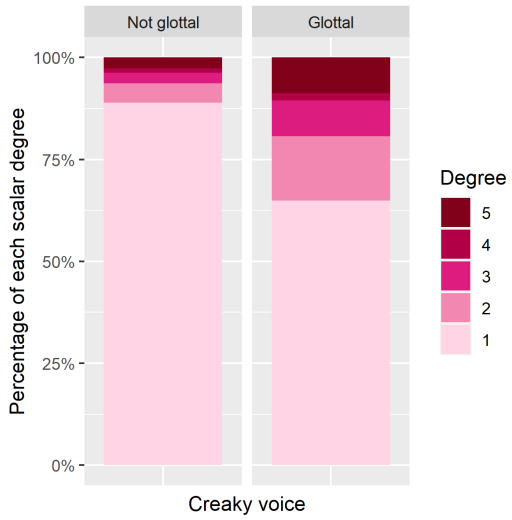


(a) Voiced stretches containing or adjacent to a glottal

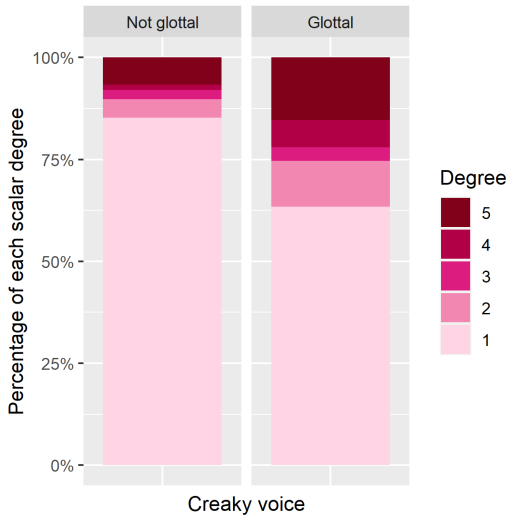


(b) Voiced stretches not containing and not adjacent to a glottal

Figure 6.21: Euler plots showing the distribution of voice quality categories and how they combine in glottal and non-glottal context

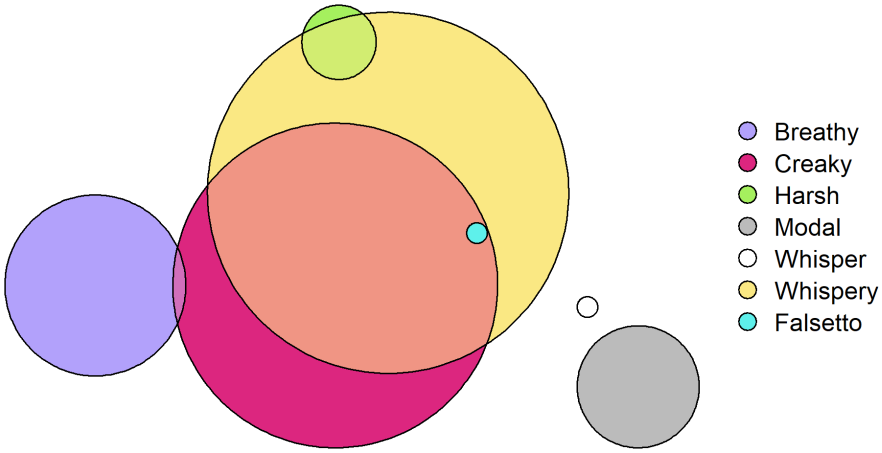


(a) Proportion of each scalar degree of creaky voice for exclusively creaky voiced stretches by glottal context

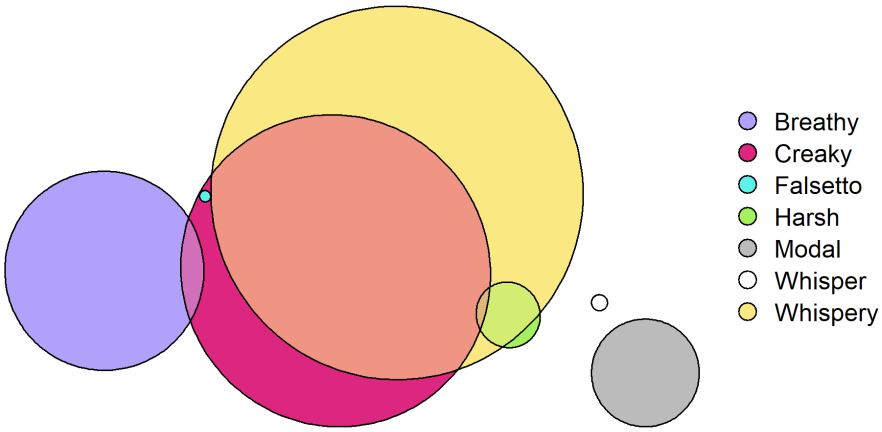


(b) Proportion of each scalar degree of creaky voice for whispery creaky voiced stretches by glottal context

Figure 6.22: Stacked percentage bar plots showing the proportion of each scalar degree of creaky voice for voiced stretches coded as exclusively creaky voice and whispery creaky voice by glottal context.



(a) Voiced stretches in phrase final position



(b) Voiced stretches not in phrase final position

Figure 6.23: Euler plots showing the distribution of voice quality categories and how they combine in final and non-final position

6.4 Discussion

I conducted a Phonation Profile Analysis of 24 speakers across Glasgow, Lothian and Shetland, including two male and two female speakers from each age group from each area, as a first step to examine how voice quality varies by gender, age, and accent in Scotland. This auditory-perceptual approach enables findings to be connected to previous VPA research on voice quality in Scotland and enables comparison between acoustic and auditory-perceptual analysis, ultimately facilitating a larger acoustic corpus study of voice quality variation by gender, age and accent in Scotland. Here, I begin by discussing the results of the PPA, followed by discussing the method.

6.4.1 Discussion of results

I predicted that whispery voice would be prevalent in the data, and that different accents might be characterised by certain components, such as harsh voice in Edinburgh (Esling 1978b) and tenseness in Glasgow (Stuart-Smith 1999a). I found that Scottish accents were characterised by whispery and tense whispery voice, but that voice quality also appears to vary between different regional accents within Scotland. Though speakers across all areas used tense, whispery, tense whispery, creaky, breathy and modal voice, each area was characterised by an increased usage of some of these qualities. Glasgow appears to be characterised by the use of tense whispery voice, consistent with the findings of Stuart-Smith (1999a). This is promising for the validity of PPA, as it suggests that PPA labels and my own rating is roughly consistent with VPA. However, the consistency of Glasgow compared to the inconsistent findings of Edinburgh might reflect Jane Stuart-Smith's involvement in the present study; I was able to confirm that my own ratings of tense and whispery voice agreed with hers, while this was not the case with the authors of other previous research who were not part of the supervisory team.

My findings in Lothian were inconsistent with the findings of Esling (1978b). Though Lothian speakers used more harsh voice than speakers from other areas, Lothian voice quality is characterised more by increased use of modal and near-modal voice qualities, as well as breathy voice. This could be partly explained by the 42 year gap between Esling's data collection in 1974 (Esling 1978b: 56) and SCOSYA data collection in 2016, allowing for Lothian voice quality to potentially become less harsh over time. This is supported by harsh voice occurring more often in older speakers and male speakers; it was adult male speakers in Esling's study (aged 22-76) who were particularly harsh, rather than boys, and the vast majority of these speakers would have been over 65 in 2016. It is possible that voice quality in Edinburgh has therefore become less harsh over time.

Sampling differences may also contribute. Esling only considered male speakers, and the inclusion of female speakers, who were less harsh here, decreases the prevalence of harshness in Lothian more generally. Furthermore, Esling found social class differentiation, with speakers from higher socio-economic class backgrounds using more creak, and speaker from lower socio-economic class backgrounds using more harsh voice. In SCOSYA, social class background was indeterminate, especially as not every participant met the criteria of not having attending higher education. The lower socio-economic status speakers who used harsh voice in Esling's study may not be represented in this sample.

Methodological differences may also be a factor. PPA shows intermittent use of harsh voice when it appears, and it is unclear at what point intermittent harsh voice is comparable to speaker-characterising harsh voice identified in Esling's work. Furthermore, Esling did not have a category for laryngeal tension in his version of VPA: What I coded as 'tense' might have been subsumed into creaky or harsh voice in Esling's scheme. Similarly, what I coded as lax and breathy voice in Edinburgh might be comparable to Esling's whispery voice.

Shetland appears to use voice qualities that appear in both Glasgow and Lothian, but in comparison uses more of scalar degrees 2-5 creaky voice, particularly scalar degree 2. Voice quality has not been studied systematically in Shetland in previous research, so this is a novel finding, and further acoustic analysis at larger scale will be useful to consider whether creaky voice is a distinctive feature of Insular Scots.

I expected that prevalence and degree of whispery voice, creaky voice and harsh voice may vary by gender. I found some similarities in voice quality between male and female speakers. Male and female speakers both used high amounts of whispery, tense, and tense whispery voice. However, male speakers used more breathy voice, modal voice and harsh voice qualities than female speakers. Male speakers also favoured different scalar degrees of different voice qualities: higher scalar degrees of breathy voice and whispery voice, and different scalar degrees of creaky voice, potentially indicating that different types of creaky voice could index gender in Scotland. Male speakers also appeared to use lax creaky voice, which did not occur for female speakers. On the other hand, female speakers were characterised mostly by using whispery, tense, and tense whispery voice, and lower scalar degrees of whispery and breathy voice. Overall, there appeared to be a general tendency for female speakers to make use of tenser voice qualities across all combination types, while male speakers seemed to adopt a less tense phonatory setting that allows use of breathy voice. Acoustic investigation of this in more speakers will be interesting as this contradicts claims of a universal tendency towards breathier voice qualities in female speakers.

Male speakers using more harsh voice is consistent with Esling (1978b) finding

that harsh voice was prevalent in a sample of only male speakers, as well as with Beck & Schaeffler (2015) finding that male adolescents use more harsh voice than female adolescents. The presence of this in a sample of adults suggests that Beck & Schaeffler's (2015) finding is unlikely to be entirely attributable to laryngeal instability during puberty in Beck & Schaeffler's (2015).

Variation by gender for other phonation types is somewhat inconsistent with Beck (1988), Stuart-Smith (1999b) and Beck & Schaeffler (2015). Contrary to Beck (1988) and Beck & Schaeffler (2015), I did find differences in terms of whispery voice, but these do not match the patterns that Stuart-Smith (1999b) identifies exactly. This is likely due to methodological differences between VPA and PPA: The findings here indicate that female speakers use whispery voice *more often*, but that when male speakers use whispery voice they tend to be *more whispery*, and VPA does not allow for patterns of this kind to be uncovered.

I also expected to find differences by age, with increased prevalence and scalar degrees of harsh, breathy and whispery voice among older speakers. The phonation types used were fairly similar between older and younger speakers, indicating that accent differences in voice quality may be relatively stable. Older speakers, however, did appear to use more harsh voice and less modal voice, which may relate to the physiological effects of ageing.

Younger speakers used more creak (scalar degree 2-5 across all combination types) than older speakers. As Dallaston & Docherty (2020) find, the claim that young female American women use creak more than other groups is widespread, but has not been empirically confirmed because most studies tend to focus largely on younger speakers. The present study is rare in comparing age groups, and provides some tentative evidence that in Scotland, younger speakers creak more than older speakers. This will need to be investigated further in the acoustic analysis in the larger sample.

As expected, creak was more prevalent in proximity to a glottal stop, and this context also favoured higher degrees of creaky voice. This demonstrates the use of PPA for coding variation in voice quality according to linguistic factors. Though there was increased creak and tense voice found in phrase-final position, this was very small; it remains to be seen whether this will be present more clearly in the acoustic analysis, which would suggest that the small difference between non-final and final contexts is an effect of the PPA coding system.

Variation by age, gender and area suggests that further acoustic study would be beneficial for understanding how voice quality varies according to social factors. With only 24 speakers, it is already possible to see some differences emerging, but the small number of speakers in each cell means that idiosyncratic variation cannot be differentiated from group-level patterns. Further acoustic study would validate the findings

of the PPA and enable interactions to be considered. The two chapters that follow investigate the relationship between PPA coding of creaky voice and automatic coding of creak using an f0-based method, and PPA coding of non-creaky voice quality and multiple acoustic measures, allowing a larger acoustic investigation of voice quality to be carried out in Chapter 9.

6.5 Reflecting on PPA

PPA showed promise in allowing laryngeal voice quality to be characterised in terms of how short-term variation contributes to overall quality. This allowed speakers who used a phonation type more frequently to be differentiated from those who used an increased auditory degree of that same quality. Thus, we can say that female speakers made more frequent use of whispery voice, while male speakers used it **less**, but were *more whispery* when they used it. We can also see how different phonation types combine, allowing insight into how speakers from Glasgow use a tense setting that accompanies their whispery voice, and we can differentiate between laryngeal tension and creak, showing that speakers from Glasgow are more tense, while speakers from Shetland use more creak. PPA also successfully identified variation according to linguistic factors. This suggests that in future research, PPA could be used to quantify variation in voice quality according to momentary conversational constraints in a way that is not possible in VPA; combining PPA with Conversation Analysis could therefore be an interesting avenue for future research.

However, certain issues with PPA also became apparent during the course of coding and analysis of the results. Here I discuss the issues of perceptually separating a voice into its components, coder reliability, and the unit of the voiced stretch.

6.5.1 Tension between componential analysis and holistic perception

As Beck (2005: 293) discusses, VPA functions as a componential system where the rater identifies the consistent components that contribute to the perception of overall quality. She notes that it has a wider scope than many other rating systems, leading to a trade-off between the number of parameters included and ease of use. As PPA draws on this componential approach, it inherits some of these difficulties; while the componential nature of VPA and PPA allow them to be comprehensive and function with reference to general phonetic theory, the number of parameters needing to be rated simultaneously in PPA introduces concerns about its reliability.

As discussed in Section 4.2.1.2, Kreiman, Gerratt & Antoñanzas-Barroso (2007) argue that listeners hear voices as a holistic pattern, leading auditory-perceptual analysis to be challenging when it requires breaking voices down into their constituent components. Later, Kreiman & Gerratt (2018: 270) imply that VPA is limited because it does not specify how different components interact. Research on VPA suggests that different settings do interact: Segundo et al. (2018) used cluster analysis to identify groups of speakers with similar VPA profiles in a study where three raters rated 99 SSBE male speakers, and found one cluster was characterised by lowered larynx, creaky voice and breathy voice, while a second cluster included speakers who used raised larynx, harsh voice, tense larynx and whispery voice. San Segundo et al. (2019) investigated correlations between different VPA settings in this same data and found that raised larynx, tense larynx, and harsh phonation were correlated with each other. However, San Segundo et al. (2019) argue that these relationships are likely the result of physiological inter-dependence and sociolinguistic patterning, rather than listeners being unable to differentiate settings, because raters do rate some speakers independently on these settings. Similar relationships between settings also seem to be revealed in the present analysis, with tense voice and whispery voice co-occurring regularly.

San Segundo et al. (2019) presented the results of a calibration session, where the three analysis discussed where their rating diverged after rating 10 voices. They identified a number of difficulties in implementing VPA, two of which I also encountered in my research: Implementing scalar ratings, and distinguishing breathy and whispery voice.

As explored by Kreiman, Gerratt & Antoñanzas-Barroso (2007), listeners have difficulty rating components of voice quality on ordinal scales. I encountered this issue when rating multiple components of voice quality in the same voiced stretch. I was usually fairly certain of my judgements where they involved an additional '1' for breathy or creaky, signifying a degree of laxness or tenseness, but rating multiple components was more challenging elsewhere. I often re-visited cases of creaky voice that I had initially rated exclusively creaky to add a scalar degree of whispery voice.

Rarer phonation types and combinations were more difficult to rate on scalar degrees, as occurring less frequently meant that there were fewer samples for comparison. I suspect lower internal consistency for settings such as lax creak than for a setting like whispery voice, which occurred frequently. It also felt easier to rate more extreme deviations from modal voice, which warranted ratings of higher scalar degrees, than cases that were closer to modal.

Like the raters in San Segundo et al. (2019), I also found distinguishing between cases of breathy voice and whispery voice difficult, although my confidence in my ability to distinguish them improved with practice and after confirming the breathy quality

of certain samples with my supervisors. Acoustic analysis of PPA-rated breathy and whispery voice, detailed in Chapter 8, will be useful in investigating whether these two phonation types were differentiated reliably.

6.5.2 Coder reliability

Within the limits of this thesis, I acted as the sole coder, though as noted in Section 6.2 the protocol and difficult issues were discussed with two VPA-trained phoneticians. However, previous research on VPA has demonstrated that different raters do not always consistently agree on VPA ratings, with San Segundo et al. (2019) finding 87% agreement between the three raters for creaky voice and 83% for breathy voice. Beck (2005) raises some concerns about intra- and inter-rater reliability when rating phonation in shorter stretches, citing Vieira (1997) who find poor interjudge and intrajudge agreement for phonation settings when analysis was based on isolated vowels. Future research may wish to work with multiple coders, and investigate how intra- and inter-rater reliability compare between VPA and PPA. It would also be beneficial to work towards developing a practical way to present stimuli in random order, which has previously been used in single-coder VPA research (Stuart-Smith 1999b). This would reduce the likelihood of predictions that the researcher has about how voice quality might pattern influencing the findings of the study. A limitation of the present study is that I completed this analysis having read VPA research on Scottish accents by Stuart-Smith (1999a) and Esling et al. (2019), upon which I based predictions about what patterns I might find. While I hope that this did not affect my coding, in the absence of a second rater, the effect of this on my ratings is currently unknown.

6.5.3 Length and segmental composition of the voiced stretch unit

In this study, I used the stretches of sonorants of 100 ms or longer as a unit of analysis. However, these stretches could vary both in length and segmental composition, which raises some potential issues.

Firstly, as Lotto, Holt & Kluender (1997) explored, voice quality in vowels also affects how vowel quality is perceived. They compared how listeners identified vowel quality in synthesised breathy and modal vowels, and found that listeners were more likely to identify higher vowels in tokens with breathy quality. If voice quality can affect the perception of vowel quality, it follows that the inverse of this (that vowel quality may affect the perception of voice quality) might also be true, and may have affected judgements in this research. In PPA, some stretches contained a single vowel, while others contained a number of consecutive segments, which could be vowels with

differing qualities or a number of sonorant consonants. Reflecting on my experience conducting PPA, I found comparing quality between stretches that different in phonemic content difficult, and it is currently unclear whether this difficulty affected my ratings. The potential influence of segmental composition on voice quality ratings should be investigated further in future research.

As Simpson (2012) illustrates, the location of nasal poles is dependent on f0, which causes nasalisation to have a similar acoustic manifestation to breathy voice in speakers with higher f0. It follows that the presence or absence of nasal poles could make auditory identification of different voice quality components more difficult. In my experience rating with PPA, I found comparison between stretches that contained nasals and stretches that did not to be particularly difficult; whether this difficulty translated to a difference in my ability to differentiate nasality and breathy voice is currently unknown.

I also found limitations in the length of the voiced stretch as a unit of analysis. Firstly, small fluctuations in the degree of a certain voice quality went unrecorded if they occurred within a voiced stretch, which limits PPA's ability to fulfill its aim of capturing short-term variation. Secondly, at the analysis stage this meant that the length of the stretch was not accounted for: a 100 ms stretch and a 300 ms stretch were ultimately taken as having an equal contribution to voice quality, despite the fact that 300 ms stretches were three times as long.

6.5.4 Issues with implementing PPA and recommendations for future users

One issue with PPA is that auditory-perceptual analysis requires data to be prepared before it is analysed, and in PPA these stages have an ability to influence each other. To prepare data for analysis, PPA involves force-aligning data and then using this to select voiced stretches, which also involves correcting alignment, and excluding portions with excessive background noise.

Initially, I did not separate preparing data from PPA analysis. This led to me encountering difficulty switching between these different steps, and paying attention to each of these things simultaneously meant that I sometimes entered scalar degrees on the wrong tier, or missed a voiced stretch entirely. I also began to worry that hearing the presence or absence of a particular voice quality was beginning to influence my alignment correction: I knew that 100 ms was my cut-off for the minimum length of a voiced stretch, and when I found a case that was almost 100 ms but not quite, I sometimes found myself tempted to adjust the boundaries so that the stretch could be included. Because of these issues, I changed how I approached the coding process

and separated data preparation from data analysis, but this did not completely resolve the issue. I sometimes still encountered places where the forced alignment needed correcting which I had missed the first time around. Ideally, these two stages would be completely separated, with one researcher preparing data and another coding for voice quality; in this case, if the voice quality coder encountered a instance where they thought the alignment needed adjusting, they could flag this for the other coder to return to, minimising the chance of the two stages influencing each other.

PPA requires a high amount of concentration from the coder to be able to identify the components present in a stretch and rate them on scalar degrees. While this is issue is not unique to PPA and is something that other auditory-perceptual coding schemes also face, I believe my own set up when conducting PPA was not conducive to maximal concentration. I often needed to open another window to compare the current voice to a previously rated sample, and then had difficulty finding my place when I returned the original sample. In future research, I would recommend that the coder uses a dual-screen set-up while coding, to mitigate the difficulties of comparing the voice to other samples.

Overall, PPA allowed investigation of how short-term variation in phonation types compounds to produce variation in overall quality according to social factors. However, auditory-perceptual analysis is limited in terms of the scale that it can be conducted at. In order to investigate variation in voice quality in Scottish accents on a larger scale, I plan to conduct an acoustic analysis, which is presented in Chapter 9. In order to conduct this analysis with reference to the auditory quality of voice quality in Scotland, I turn now to an investigation of how PPA relates to automated f0-based coding of creaky voice and multi-measure acoustic analysis of non-creaky voice quality.

Chapter 7

Connecting PPA and f0-based analysis of creaky voice

7.1 Introduction

In Chapter 6, I conducted a Phonation Profile Analysis of 24 speakers stratified by age (12 older 65+, 12 younger 18-25), gender (12 male, 12 female) and area (8 Glasgow, 8 Lothian, 8 Shetland), and found that Shetlanders used more creak that speakers from other areas, that younger speakers used more creak than older speakers, and that glottal stops favoured creak. In Chapter 9, I will consider whether these patterns persist in an acoustic analysis of more data and speakers from SCOSYA. However, tense voice and creak may be difficult to differentiate on the basis of acoustic measures alone, as they both involve laryngeal constriction. Because of this, I turned to automatic methods of identifying creak as a potential way of separating creak from non-creak before proceeding with acoustic analysis using other acoustic measures. In this chapter, I consider how automatic identification of creak relates to creak coded using PPA, and consider whether this might be a viable option for quantifying creak in the larger corpus and aid the interpretation of results with reference to auditory-perceptual descriptive labels. In this chapter, the terms 'creak' and 'creaky voice' refer to only types of creak that involve alterations to F0, and exclude tense voice.

In Section 4.2.5.2, I outlined the existence of a number of automatic methods of detecting creak and summarised existing research on their effectiveness. In designing the present study, I trialled implementation of automatic detection of creak alongside acoustic analysis of non-creaky voice. I aimed to consider the possibility of whether using these approaches in tandem might allow for a more straightforward interpretation of acoustic measures in acoustic analysis of non-creaky voice.

In the exploratory analysis presented here, I implement f0-based automatic detection of creak (Dorreen 2017, Dallaston & Docherty 2019) using REAPER (Talkin 2015) to code creaky glottal pulses. I then compare automatic creak detection to manual coding using PPA. Previous research evaluating the success of automated methods largely focuses on numerical evaluations how they compare to manually-annotated creak. Here, I instead focus more on investigating the causes of mismatches between the two methods. In doing so, I aim to proceed in implementing this method on a larger scale with a better idea of its advantages and limitations.

7.2 Methods

I used Dallaston & Docherty's (2019) f0-based method of automatically detecting creaky voice to detect creak in the PPA-coded sub-corpus and compare this with the auditory coding of creak using PPA, to evaluate the possibility of using this method in Chapter 9.

Files from the PPA-coded sub-corpus were processed using MacReaper (Dallaston & Docherty 2019), a drag-and-drop interface for REAPER (Talkin 2015). The files were processed at 16 kHz, because Talkin (2015) notes processing sound files above 16 kHz will 'incur quadratic increase in computational requirements without gaining much in output accuracy'.

MacReaper's outputs includes the f0 track, location of voicing detected, and the estimated location of Glottal Closure Instants (GCIs). To allow comparison with PPA-annotated creak, only GCIs detected within a manually-annotated PPA-coded voiced stretch were considered. Following Dallaston & Docherty (2019), local f0 was calculated for each GCI by taking the inverse of the time between each GCI that occurred within a voiced stretch. Local f0 was then used to create a distribution of GCI f0. Antimodes were then detected using the automated procedure described in Dallaston & Docherty (2019) using an R script (R Core Team 2020) obtained via personal communication with Katherine Dallaston. The script uses the `modes` package (Deevi & Strategies 2016) to identify the f0 mode of non-creaky speech, the f0 mode of creak, and the antimode between them using the following procedure:

1. Find the maximum peak in the distribution, assumed to be the mode of a speaker's non-creaky f0
2. Find other peaks in the distribution below the f0 of the maximum peak
3. Test each peak below the maximum peak, starting at the one closest to it, and look for an antimode with a density of <0.005 between them
4. Repeat this until an antimode with a sufficiently low density is found

Using this procedure, any GCI with an f0 below a speaker's antimode was coded as creaky, and any GCI with an f0 above a speaker's antimode was coded as non-creaky. The automated procedure also allows an antimode to be set to a default value if no antimode is found in the distribution, but this was not applied here in order to consider cases where no antimode could be found in more detail.

To preview the results, visual inspection of the f0 distributions and identified modes and antimodes suggested that there was some difficulty in identifying an f0 antimode

and/or creaky mode for 8 out of 24 speakers. I attributed this difficulty to a speaker having a unimodal f0 distribution in one case, and to the automated procedure identifying the creaky f0 as being the closest max peak in f0 below the non-creaky mode in the other 7 cases, resulting in part of the non-creaky f0 distribution being considered to be creak. I therefore altered the procedure: Instead of testing peaks below the maximum f0 peak starting with the closest peak, I tested peaks below the maximum peak in order of height. This brought automatic identification of creaky modes and f0 antimodes in line with the ones that could be identified visually for 6 out of 7 speakers.

In the results section that follows, I refer to the first version of the procedure that I tested as the 'original procedure', and the second version as the 'altered procedure'.

7.2.0.1 Comparison between PPA and f0-based coding

The output of the automated procedure was compared to the PPA coded of creak to establish the main similarities and differences between these two methods of identifying creak.

When the two approaches are compared, automated detection of creak is compared to scalar degrees 2-5 of PPA-coded creak, and excludes scalar degree 1 (tense voice), which is considered 'non-creaky'.

Two types of agreement (positive and negative) and two types of disagreement (antimode only and PPA only) were defined to allow agreement between the two systems to be quantified. These were defined as follows:

- Positive agreement: Creak present according to both PPA coding and automated procedure
- Negative agreement: Creak not present according to either method
- Antimode only: Automated procedure detects creak, but PPA does not
- PPA only: PPA coded creak is present, but automated procedure does not detect creak

PPA and REAPER operate at different units of analysis (the voiced stretch, compared to the GCI), so overall rates of agreement between the two systems were calculated both at the level of the voiced stretch and at the level of the GCI, as outlined below. Where the voiced stretch was taken as the unit of analysis, positive agreement required all GCIs to be below the speaker's antimode, while negative agreement required no creaky GCIs coded by the automated procedure. Disagreement in *any* GCI within a stretch led the whole stretch to be treated as a disagreement.

A second analysis took the GCI as the unit of analysis, allowing for a more precise comparison of the amount GCIs within a stretch where antimode coding aligned with PPA coding. Finally, agreement was considered for 10ms chunks of time, allowing comparison of F1 scores with White et al. (2022) and consideration of how agreement varied by speaker demographics.

7.2.0.2 Reasons underlying disagreements

Cases of disagreement between the two coding systems (antimode only and PPA only, taking the stretch as the unit of analysis) were considered in more detail. The spectrogram, waveform, and glottal pulses identified by MacReaper in Praat were then inspected visually and the auditory quality of the stretch was checked. This analysis resulted in seven categories of disagreement:

- Antimode error: A speaker using an f0 below their f0 antimode in non-creak, or vice-versa
- Coding error: A stretch was coded incorrectly by the researcher in PPA
- F0 tracking error: An error resulting from inaccurate detection of GCIs (commonly pitch halving and pitch doubling)
- Harsh voice at low pitch: Antimode-only creak which occurs in a section coded as harsh voice which exhibits a low f0
- Time-resolution differences: Automated procedure and PPA coding are in agreement in places, but disagreements occur on individual GCIs within stretch because the different methods define creak differently and work with different units of analysis
- Multiply pulsed creak: A particular type of creak which exhibits multiple f0s means that automatic detection of creak based on a definition of low f0 does not make sense
- Multiple errors: Different GCIs within the stretch disagree with PPA coding for different reasons

Time-resolution differences were investigated further by considering the percentage of GCIs in each voiced stretch that agreed or disagreed with the PPA coding of creak.

7.3 Results of F0-based identification of creak with REAPER

7.3.1 Speaker F0 distributions

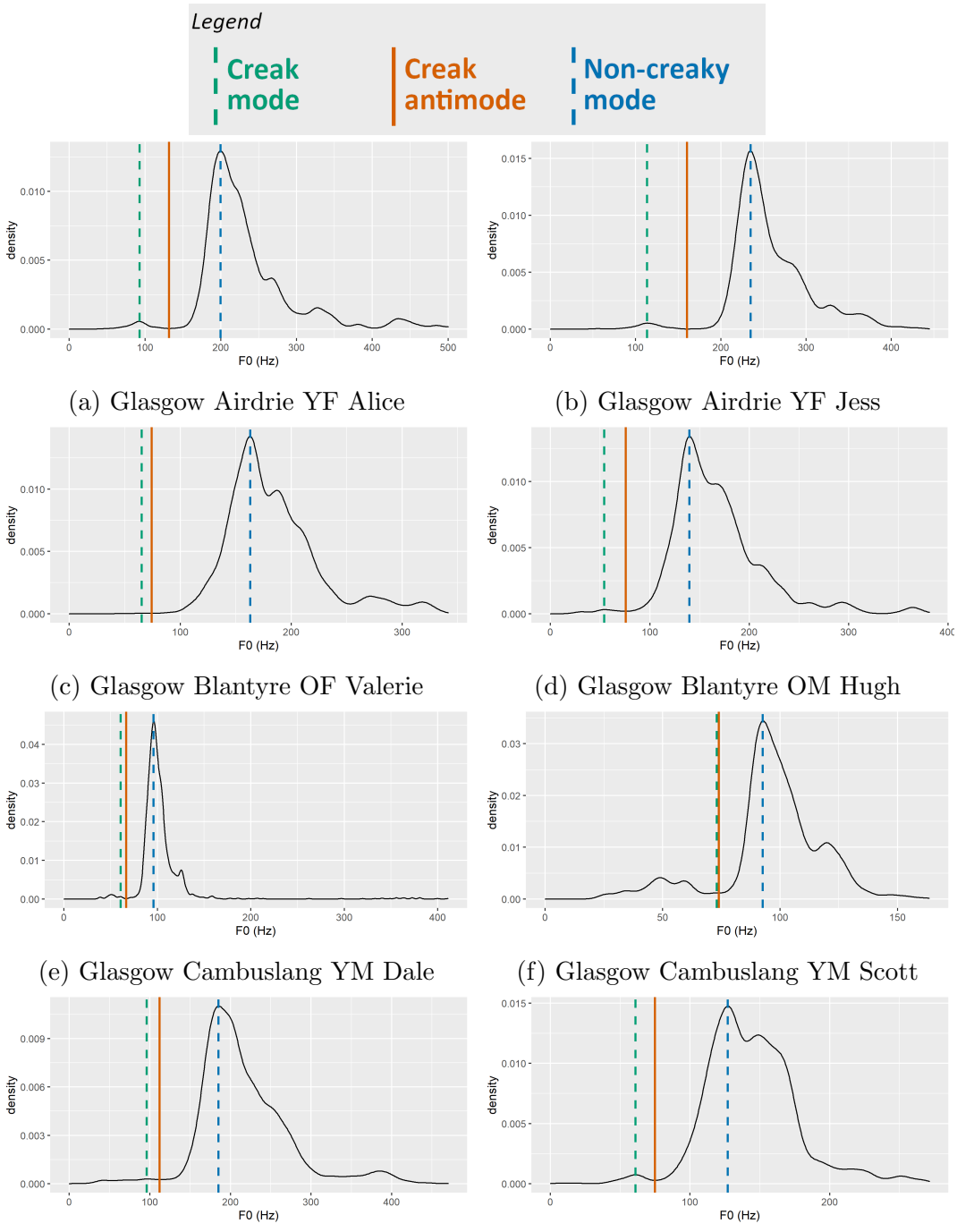
Table 7.1 shows each speaker's non-creaky f0 mode and creaky f0 mode, and the antimode between them.

Area	Age	Gender	Pseudonym	Antimode (Hz)	Non-creaky mode (Hz)	Creaky mode (Hz)
Glasgow	Y	F	Alice	132	200	93
Glasgow	Y	F	Jess	160	235	113
Glasgow	O	F	Valerie	74	163	65
Glasgow	O	M	Hugh	76	140	54
Glasgow	Y	M	Dale	67	96	61
Glasgow	Y	M	Scott	74	93	73
Glasgow	O	F	Rhona	112	185	96
Glasgow	O	M	Jack	75	127	61
Lothian	O	F	Tina	110	163	95
Lothian	O	M	Gary	69	109	51
Lothian	Y	F	Kayleigh	117	180	99
Lothian	Y	M	Toby	54	106	28
Lothian	O	F	Margaret	n/a	189	n/a
Lothian	O	M	Bruce	76	133	64
Lothian	Y	F	Grace	116	190	98
Lothian	Y	M	Finlay	75	125	57
Shetland	O	F	Betty	152	202	150
Shetland	O	F	Jane	128	188	86
Shetland	Y	F	Kellie	158	206	154
Shetland	Y	F	Stephanie	138	167	129
Shetland	Y	M	Lewis	88	121	76
Shetland	O	M	Alexander	79	117	73
Shetland	Y	M	Graham	90	113	86
Shetland	O	M	Hugh	65	94	56

Table 7.1: Non-creaky f0 modes, f0 antimodes, and creaky f0 modes by speaker. All values are given in Hz. Y = Younger (approx 18-25); O = Older (approx 65+); F = Female; M = Male.

The original automated procedure identified an f0 antimode for all speakers except from Margaret, whose antimode and creaky mode are both listed as 0. Margaret's f0 distribution is shown in Figure 7.2e. Figure 7.1, Figure 7.2 and Figure 7.3 show f0 distributions for each speaker, where the mode of non-creaky speech is marked with a dashed blue line, the antimode is marked with a red line, and the creaky mode is marked with a dashed green line.

Visual inspection of the f0 distributions suggests that the f0 mode was correctly identified for all speakers, but that the original automated procedure had some issues



(a) Glasgow Airdrie YF Alice

(b) Glasgow Airdrie YF Jess

(c) Glasgow Blantyre OF Valerie

(d) Glasgow Blantyre OM Hugh

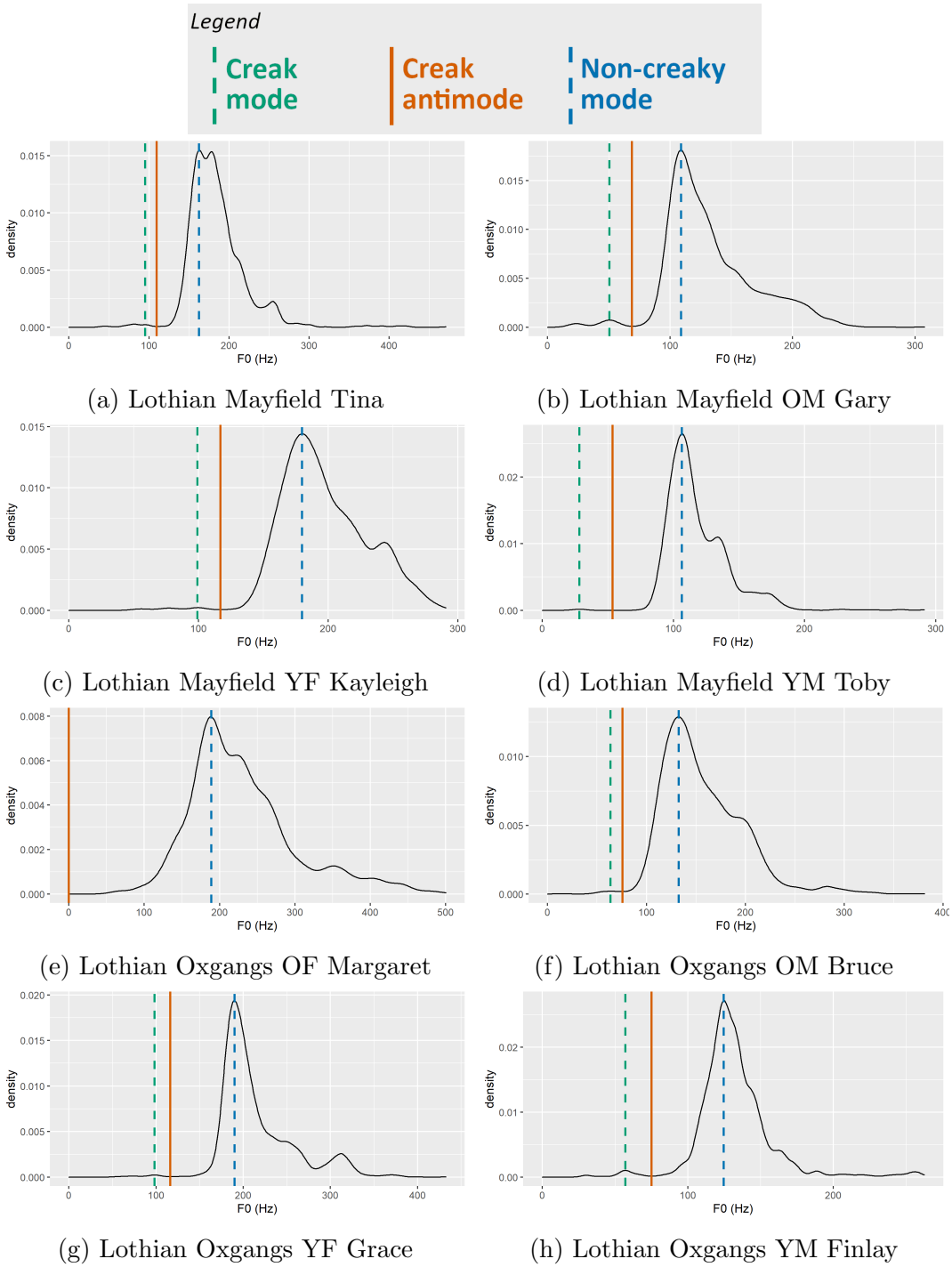
(e) Glasgow Cambuslang YM Dale

(f) Glasgow Cambuslang YM Scott

(g) Glasgow Dumbarton OF Rhona

(h) Glasgow Dumbarton OM Jack

Figure 7.1: F0 distributions for speakers from Glasgow, with modes and antimodes identified using the original procedure



(a) Lothian Mayfield Tina

(b) Lothian Mayfield OM Gary

(c) Lothian Mayfield YF Kayleigh

(d) Lothian Mayfield YM Toby

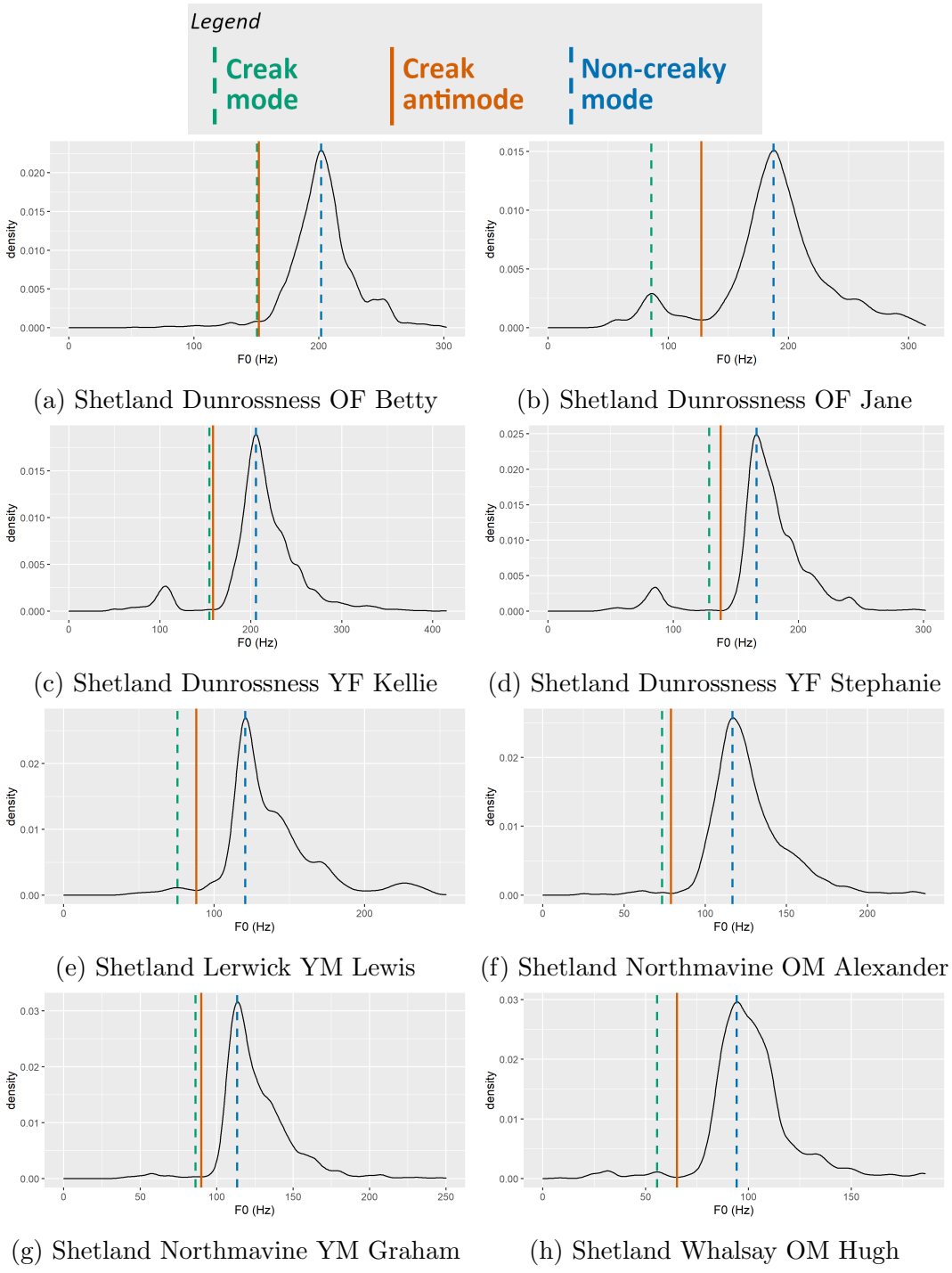
(e) Lothian Oxgangs OF Margaret

(f) Lothian Oxgangs OM Bruce

(g) Lothian Oxgangs YF Grace

(h) Lothian Oxgangs YM Finlay

Figure 7.2: F0 distributions for speakers from Lothian, with modes and antimodes identified using the original procedure



(a) Shetland Dunrossness OF Betty

(b) Shetland Dunrossness OF Jane

(c) Shetland Dunrossness YF Kellie

(d) Shetland Dunrossness YF Stephanie

(e) Shetland Lerwick YM Lewis

(f) Shetland Northmavine OM Alexander

(g) Shetland Northmavine YM Graham

(h) Shetland Whalsay OM Hugh

Figure 7.3: F0 distributions for speakers from Shetland, with modes and antimodes identified using the original procedure

identifying either the antimode or creaky mode in some speakers. Visual inspection revealed that the creaky mode was too high for Dale (YM Glasgow), Scott (YM Glasgow), Betty (OF Shetland), Kellie (YF Shetland), Alexander (OM Shetland), and Graham (YM Shetland), resulting in antimodes that were too high for Scott, Betty, Kellie and Graham. The comparison between PPA and automated identification of creak given below is based on the data from the original automated procedure, where the antimodes for these speakers are not in line with what visual inspection produces. However, Section 7.3.6 outlines the changes that resulted to antimodes and disagreements/agreements between the two coding systems when this procedure was altered as set out in Section 7.2.

7.3.2 Comparison between PPA and f0-based coding at the level of the PPA voiced stretch

2170 PPA voiced stretches were considered, but 22 were excluded because REAPER did not detect any GCIs. 2148 stretches are included in the following analysis.

F0-based coding of creak agreed with my PPA coding of creak in 87% of voiced stretches (n = 1859). Of these, there was negative agreement (creak not present according to either method) between PPA coding and f0-based coding on the absence of creak throughout the stretch in 1797 cases and positive agreement on the presence of creak throughout the stretch in 62 cases.

There was some level of disagreement between the two coding schemes in the remaining 13% of stretches (n = 289). 164 of these cases were antimode-only identification of creak, while 125 were PPA-only cases of creak.

The amount and type of disagreements varied between speakers, as shown in Figure 7.4. For example, as shown in Figure 7.2e, Margaret's f0 antimode could not be detected, so no creak was coded using the f0-based coding, leading to negative agreement in 85% of cases (n= 84) but PPA-only in the remaining 15% (n = 15) of stretches.

The highest overall agreement between PPA coding and f0-based coding was in the case of Valerie, where negative agreement (Creak not present according to either method) occurred in 98% of stretches (n = 82). However, there was also no positive agreement between the two coding systems in the coding of creaky stretches the case of Valerie, as the only two creaky-voiced stretches that were coded as creaky in PPA were missed in the f0-based coding.

This highest rate of positive agreement between PPA and f0-based coding of creak was for Scott (n = 95), where 17% of stretches (n = 16) were classed as creaky by both systems. The two coding schemes also showed negative agreement for 63% of

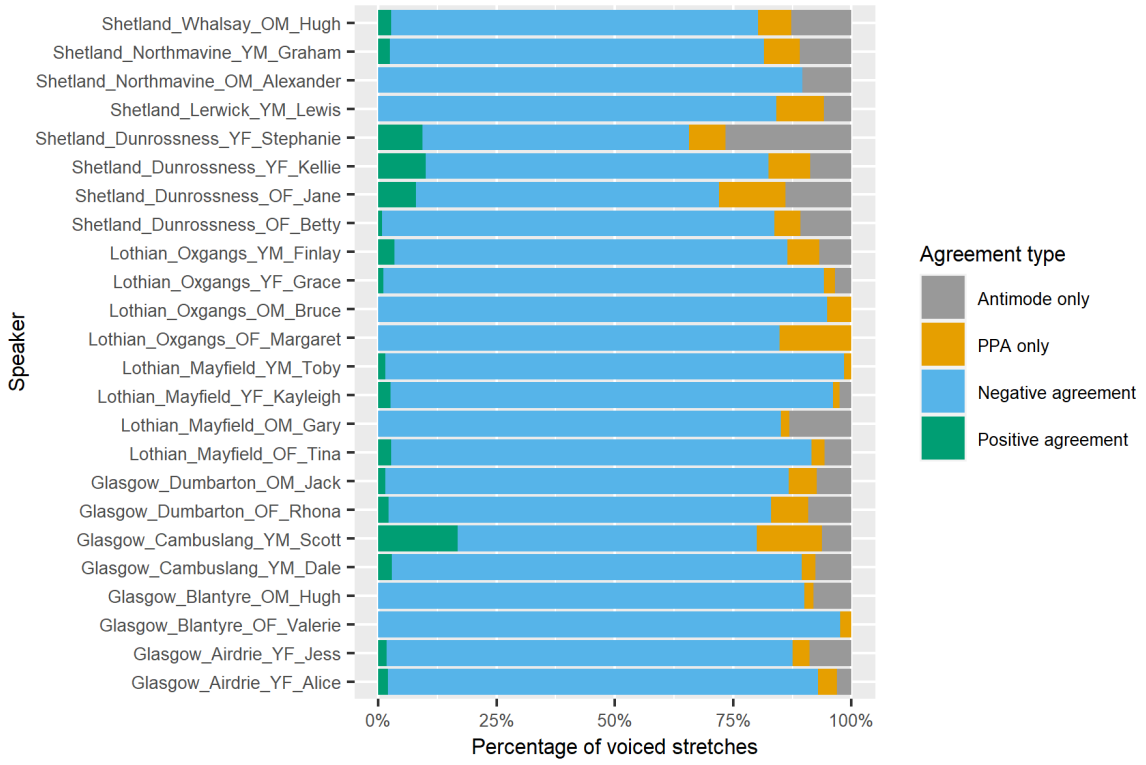


Figure 7.4: Agreement between PPA and f0-based coding of creak by speaker, with unit as the Voiced Stretch from PPA coding

non-creaky stretches (n = 60), while they also disagreed in 20% of cases, with the antimode-only creak identified in 6 stretches and PPA-only creak coded in 13 stretches.

7.3.3 Comparison between PPA and f0-based coding at the level of the GCI

65,616 GCIs were considered, of which 1,653 (3%) were creaky according to the f0-based method, and 2,504 were creaky in PPA (4%). When GCIs were taken as the unit of analysis, f0-based coding of creak agreed with my PPA coding in 97% of GCIs (n = 63,771).

Negative agreement occurred in 62,615 GCIs (95% of GCIs), where both methods agreed the pulse was not creaky, while positive agreement occurred in 1,156 GCIs (2% of GCIs), where both methods agreed the pulse was creaky. F0-based coding disagreed with my PPA coding of creak in the remaining 3% of GCIs (n = 1,845). Of these, 1,348 were PPA-only creak and 497 were antimode-only creak.

As shown in Figure 7.5, the amount and types of disagreements identified varied between speakers.

Within the 289 stretches that contained a disagreement between the two coding

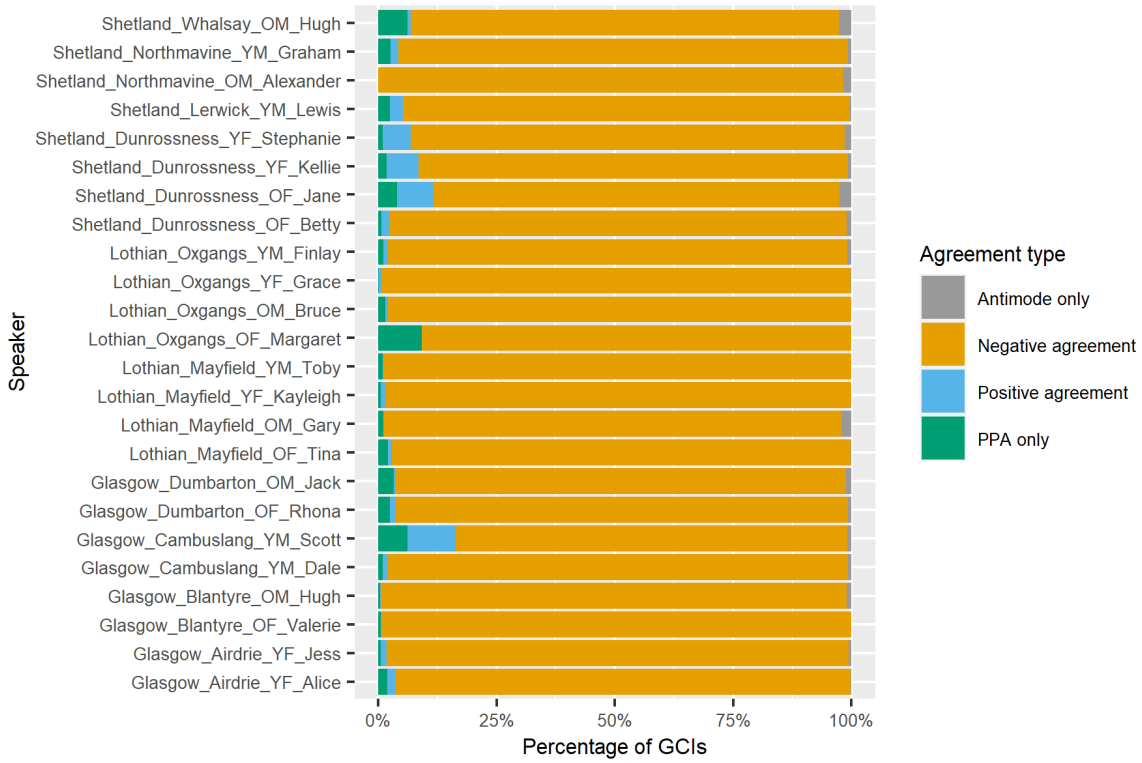


Figure 7.5: Agreement between PPA and f0-based coding of creak by speaker

% of GCIs in stretch where automated coding agrees with PPA	Number of stretches	% of
None	71	
$< 25\%$	16	
25-50%	31	
50-75%	26	
> 75 %	145	

Table 7.2: Distribution of PPA voiced stretches by level of agreement between the two coding schemes at the level of the GCI unit

schemes, 25% of those stretches (n=71) did not contain any agreement between the two different coding schemes. However, as shown in Figure 7.6, in most stretches that contained a disagreement, the f0-based coding matched the PPA-coding for at least one GCI. As summarised in Table 7.2, in many of these cases there was still a high percentage of agreement between the two coding schemes with 50% of stretches that contained a disagreement matching for at least 75% of the stretch.

7.3.4 Time-based comparison and *F1* scores

White et al. (2022) analysed data in 10ms chunks, for so the purposes of enabling comparison, I also present agreement between PPA coding and f0-based coding according by time in ms. Because this the data was not coded in 10ms chunks in either the f0-based system or in PPA, this is only presented for comparison to previous research and to allow calculation of $F1$ scores, shown below.

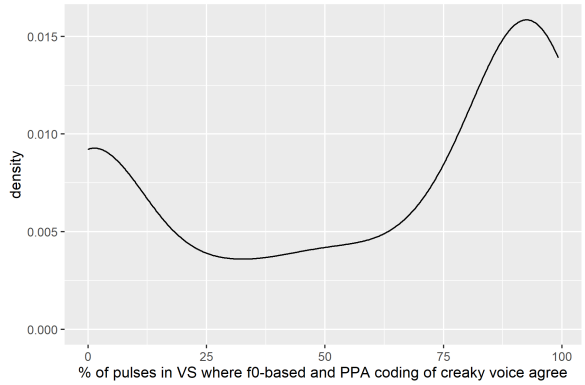


Figure 7.6: The percentage of GCIs that agree within each voiced stretch that contain at least one disagreement

Agreement type	Time (ms)	percentage
Antimode only	110.5	0.7
Negative agreement	15896.1	96.5
Positive agreement	216.0	1.3
PPA only	248.5	1.5
Total	16,471	100

To enable comparison to previous research by White et al. (2022), I calculated the $F1$ score, a method of often used to evaluate the performance of automated methods of identifying creak because it accounts for the fact that creak occurs much less often than non-creak. $F1$ score therefore discounts cases where both coding systems agree on the absence of creak, and is calculated according to the following equation.

$$F1 = \frac{2 \times \text{True Positives}}{2 \times \text{True Positives} + \text{False Positives} + \text{False Negatives}} \tag{7.1}$$

It can be roughly interpreted in terms of values closer to 1 representing better performance, and values closer to 0 representing worse performance. I this study, I do not conceptualise differences between the two coding schemes as true or false positives or negatives, but each of these has an equivalent in my own terminology, meaning that I was able to calculate $F1$ score as follows:

$$F1 = \frac{2 \times \text{Positive Agreement}}{2 \times \text{Positive Agreement} + \text{Antimode Only} + \text{PPA only}} \tag{7.2}$$

Across all speakers, $F1$ was 0.55.

F1 scores differed considerably between speaker groups. The highest $F1$ score was for young female speakers, at 0.77, followed by younger male speakers, at 0.53. Older speakers had lower $F1$ score than their younger counterparts, with $F1$ for older female speakers at 0.43 and $F1$ for male speakers at 0.17.

7.3.5 Inspection of disagreements

Stretches that contained any disagreement were inspected for potential explanation for the disagreement. This resulted in seven categories of disagreements, allowing evaluation of the frequency of each category of disagreement and how these patterned according to speaker age and gender.

Table 7.3 shows the overall counts of each type of disagreement and the percentage of disagreements that they make up. The overall proportions of error differed between antimode-only and PPA-only creak: For example, f0-tracking errors, which involve failure of REAPER to accurately track f0, resulted in far more cases where creak was identified only in the f0-based method than cases where it was identified only in PPA.

Reason	n	%
Multiply pulsed	3	1
Harsh voice	10	3
Coding error	13	4
Multiple errors	17	6
Antimode error	40	14
F0 tracking error	65	22
Time resolution	141	49
Total	289	100

Table 7.3: Distribution of different types of disagreement between the two coding schemes

Table 7.4: Distribution of different types of disagreement by gender and age

Type	OF (%)	OM (%)	YF (%)	YM (%)
Antimode error	30	8	5	8
Coding error	1	9	3	6
F0 tracking error	10	49	11	24
Harsh voice	5	9	0	0
Multiple errors	2	15	3	4
Multiply pulsed	3	0	0	0
Time resolution	48	9	78	58
Total (n)	87	65	65	72

As shown in Figure 7.8 and Table 7.4 , the type of disagreement also varied according to age and gender of the speaker. Disagreements for older male speakers more commonly involved f0 tracking errors.

In the following sections, I will present examples of each of these disagreements between the two systems. In figures like Figure 7.9, REAPER's GCIs are presented on Tier 1, with GCIs coded as creaky using the f0-based system noted as 'c'. Tier 2 then shows intervals of voicing identified by REAPER as 'phonation'. Tier 3 and 4 show the hand-corrected version of MFA's forced alignment, with Tier 3 showing

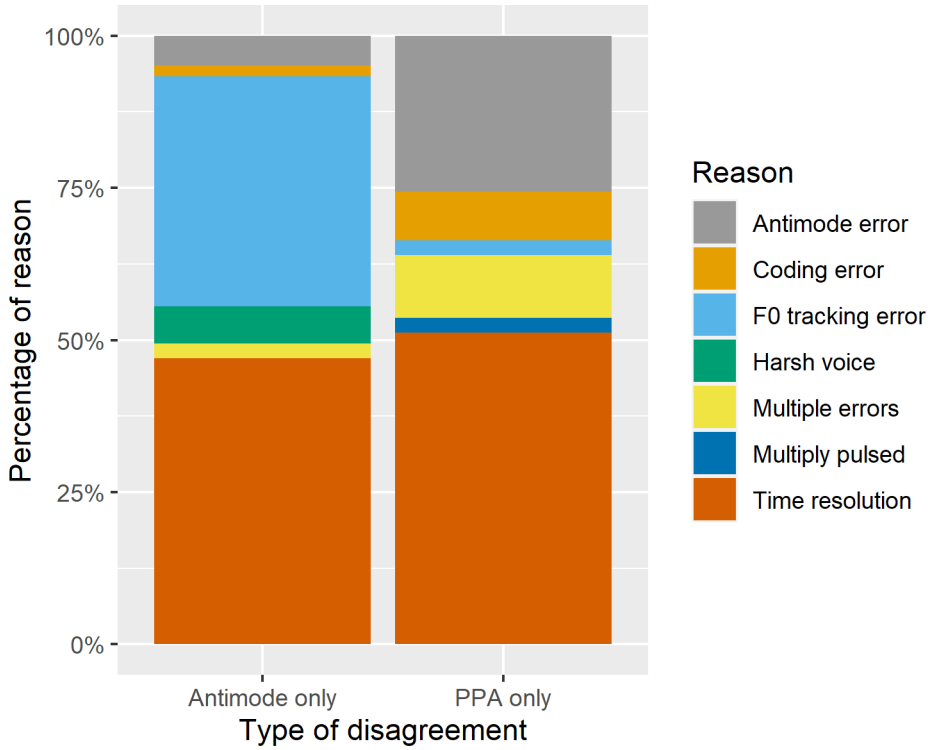


Figure 7.7: Stacked percentage bar plot showing the main disagreement types identified for discrepancies between antimode and PPA coding of creak

the orthographic transcription aligned to word-level, and Tier 4 showing the phonemic transcription, encoded in ARPABET, aligned to phone-level. Tier 5 shows glottal stops realised without a full closure. Tier 6 shows the numeric code of the stretch, in the format of 'speaker# - stretch#'.

7.3.5.1 Time resolution differences

The most common reason (n = 141, 49%) for disagreements between the two systems appeared to be differences in the time resolution of the two coding systems, which had different units of analysis. Figure 7.9 shows a case were antimode-only creak is identified GCIs towards the end of a stretch that is not creaky in PPA.

7.3.5.2 F0 tracking issues

Other disagreements related to f0 tracking. In some of cases, excluded from the numerical analysis here, REAPER did not detect any GCIs in the stretch, leading to no identification of creak.

In others, REAPER identified some GCIs in the stretch, but at the wrong frequency. Problems of this type generally occurred in cases where the voice quality in the stretch was particularly noisy, as in the case of the example shown in Figure 7.10a, where the

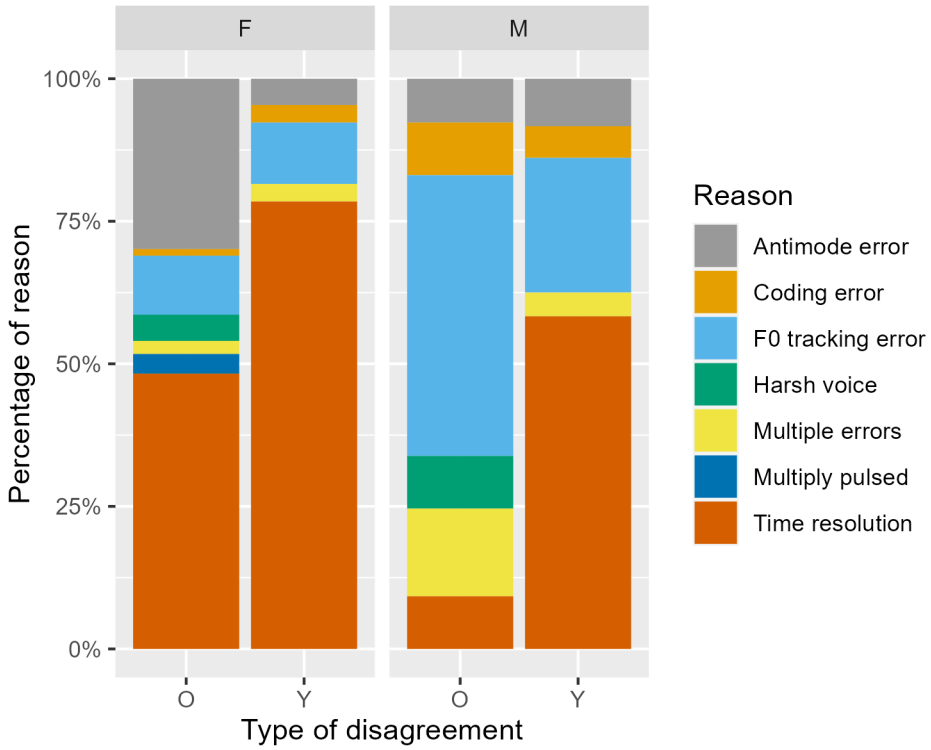


Figure 7.8: Stacked percentage bar plot showing the main disagreement types identified for discrepancies between antimode and PPA coding of creak

stretch was rated as scalar degree 5 for whispery voice.

Some cases also demonstrated pitch doubling or pitch halving. Figure 7.10b shows an example of pitch halving occuring, potentially related to the drop in amplitude that occurs towards the end of the stretch.

Low and irregular f0 found in creak sometimes led to f0 tracking issues in REAPER that did not translate to discrepencies between antimode and PPA coding. For example, as shown in Figure 7.10c, a creaky stretch was sometimes coded correctly as creaky by the f0-based method, as REAPER failed to identify each GCI, resulting in f0 below the speaker antimode.

In rare cases, REAPER identified GCIs in voicelessness. In the case in Figure 7.10d, REAPER identified GCIs in whisper at an f0 below the speaker's antimode, leading it to be coded as creak.

7.3.5.3 Problems with antimode

Furthermore, some issues stemmed from the automated procedure which identified speaker's f0 antimodes. In some of these cases, creak was identified auditorily and I confirmed the presence of it upon rechecking the case, but it occurred above the speaker's f0 antimode. Figure 7.11a shows this for Gary (Lothian OM), whose use of

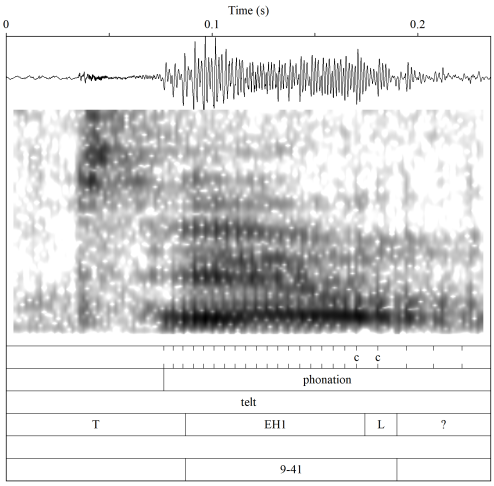


Figure 7.9: An example of where time resolution differences resulted in GCIs were annotated as creaky by the f0-based method in a voiced stretch that was not coded as creaky in PPA, in the speech of Tina, an older female speaker from Shetland.

breathy voice and harsh voice often cause pitch halving errors, leading to an artificially lowered antimode and the procedure missing instances of creak.

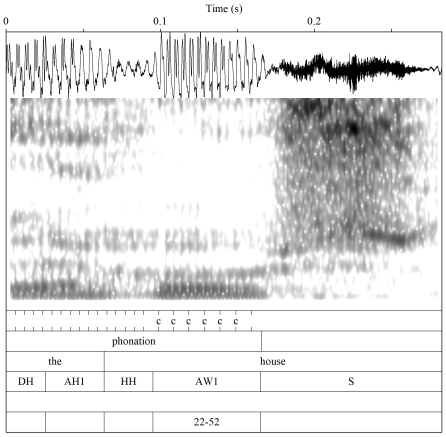
Elsewhere, this issue relates to f0 irregularity in creak. In the example shown in Figure 7.11b, creak was identified correctly by REAPER for most GCIs, except for two pulses. These pulses are close in f0 to the f0 of other pulses in the stretch, at 152 Hz and 158 Hz respectively, but don't quite make it below Betty's f0 antimode of 152 Hz.

However, Betty also exhibits non-creaky voicing below her antimode, as shown in Figure 7.11c, where her f0 extends down to 130 Hz without taking on an auditorily creaky quality, so this is not a case where a lower f0 antimode would help to identify more creak correctly overall.

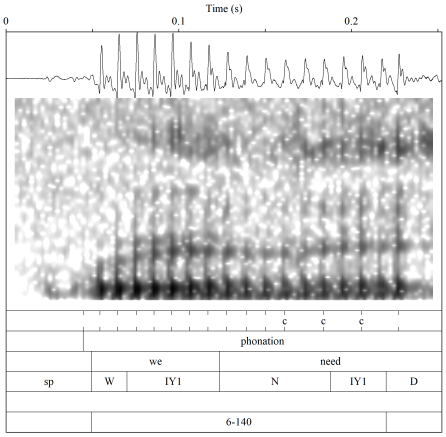
In the case of Margaret, an older female speaker from Lothian, these cases occurred because the automated procedure had not identified an antimode. Visual inspection of the waveform and spectrogram of creak produced by Margaret, an example of which is shown in Figure 7.11d, revealed a potential reason for this: In Margaret's voice, creaky voicing tends to develop over the course of a voiced stretch, and f0 lowers gradually, so that there is no clear division in her voice between creaky and non-creaky f0. This is reflected in her unimodal overall f0 distribution, shown in Figure 7.2e.

7.3.5.4 PPA coding errors

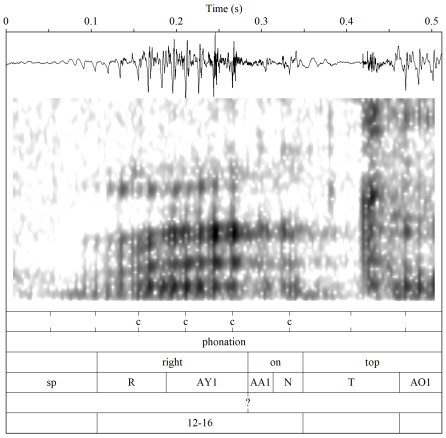
Upon review, I discovered that 4% of stretches that contained a disagreement were cases where I had incorrectly coded a stretch as creaky. These were likely due to typos or my own inexperience with auditory coding of creak in the early files coded.



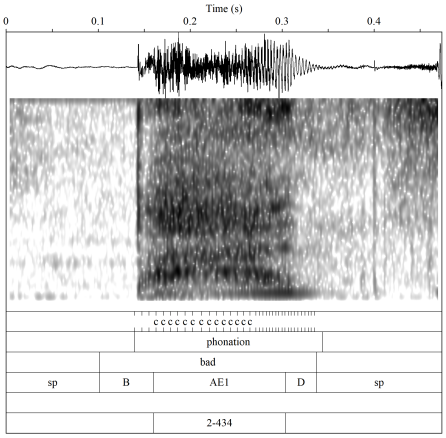
(a) REAPER demonstrating difficulty tracking f0 in highly whispery voice (Glasgow OM Alexander)



(b) Pitch halving error (Glasgow YM Scott)

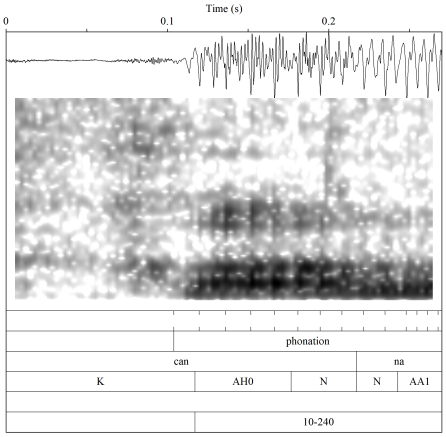


(c) Creak correctly identified in case of f0 tracking error (Lothian YM Toby)

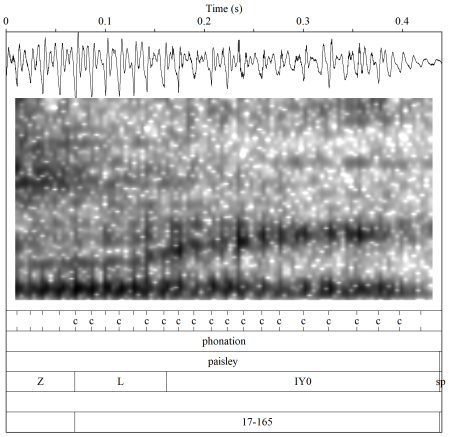


(d) An example of REAPER detecting GCIs during voiceless whisper (Glasgow YF Jess)

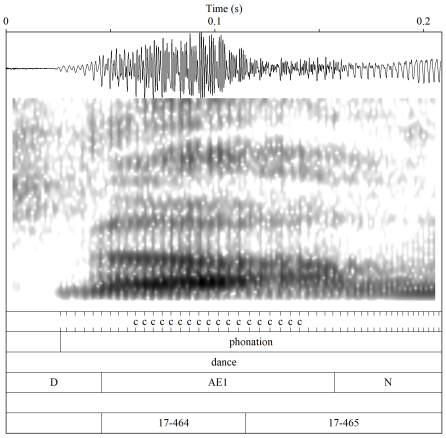
Figure 7.10: Examples of f0-tracking errors demonstrated by REAPER



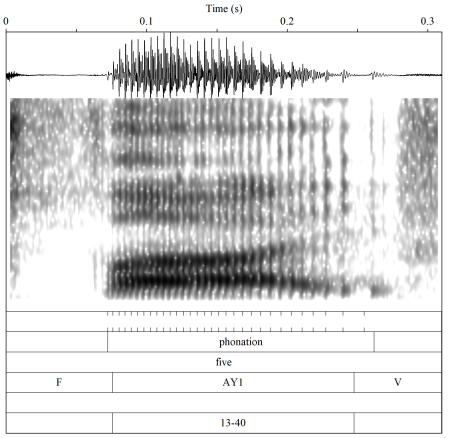
(a) Creak occurring above speaker's antimode (Lothian OM Gary)



(b) Creak occurring above speaker's antimode (Shetland OF Betty)



(c) Non-creaky voice occurring below speaker's antimode (Shetland OF Betty)



(d) A smooth transition between non-creaky and creaky f0s (Lothian OF Margaret)

Figure 7.11: Examples of antimode errors

7.3.5.5 Harsh voice

In some cases, the f0-based method annotated not creak, but harsh voice, which may involve low pitch resulting from aryepiglottic fold vibration (Esling et al. 2019: 73-75). Additionally, the aperiodic noise in harsh voice can lead to f0 tracking errors that can lead it to be coded as creak. An example of this is shown in Figure 7.12a.

7.3.5.6 Multiple pulsing

The f0-based method also struggled with instances of multiply pulsed creak, a type of creak that involve alternation between longer and shorter pulses (Keating, Garellek & Kreiman 2015: 2). In some of these cases, REAPER tracked just one of the f0s that were present, leading to creak being correctly identified. However, in other cases, this caused problems. In some instances, REAPER correctly identified both f0s, which led to the f0 being above the speaker's antimode despite being creaky, and in other cases, multiple pulsing led to other f0 tracking errors. In the example shown in Figure 7.12b, patterns in the waveform and couplets of vertical striations can be seen in the spectrogram, suggesting multiple pulsing, but REAPER also identifies additional GCIs.

Multiple pulsing also occurred in some instances identified as harsh voice in PPA. In the example shown in Figure 7.12c, REAPER tracks only one of the two possible f0 tracks, leading the f0 to fall below this speaker's antimode.

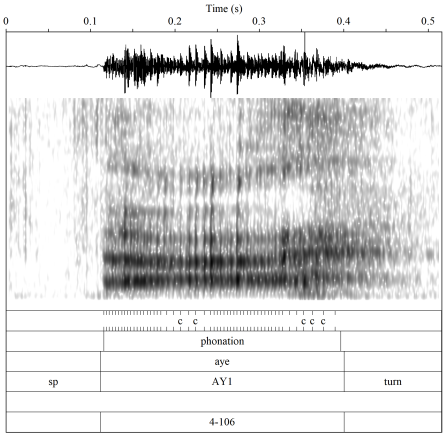
7.3.6 Re-analysis using new procedure

7.3.6.1 Changes to antimodes and creaky modes

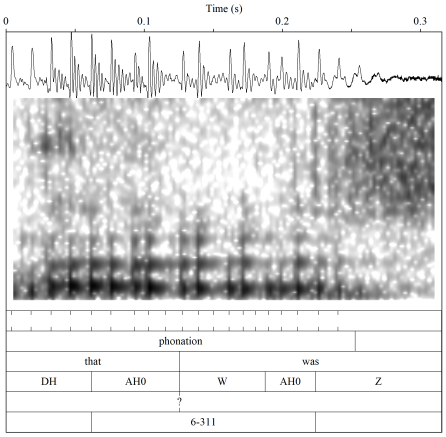
Because of the discrepancies between creaky modes and antimodes identified by the automated procedure, and those that could be identified with a visual check, I altered the original automated procedure as described in Section 7.2.

To recap, the original automated procedure involved finding the maximum peak in the f0 distribution (assumed to be the non-creaky antimode), then testing lower f0 peaks to find which was separated from the maximum peak by a sufficiently low antimode, starting with the peak *closest* in f0. The new automated procedure tested each peak in order of height, rather than in order of proximity to the non-creaky mode.

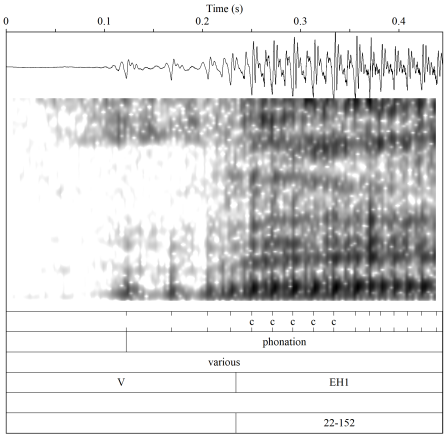
I re-ran the analysis using the new procedure for all 24 speakers. As shown in Table 7.5, the new procedure identified new creaky modes new creaky modes for 6 out of the 7 of the speakers whose creaky modes looked to be incorrectly identified in the original



(a) Harsh voice (Glasgow OM Hugh)



(b) Multiply pulsed creak (Glasgow YM Scott)



(c) Harsh voice with multiple pulsing (Shetland OM Alexander)

Figure 7.12: Examples of harsh voice and multiple pulsing

procedure. The new procedure also identified new creaky modes for two speaker for whom I had not identified any issues (Lothian OF Tina and Shetland OM Hugh).

In three out of the four speakers whose antimodes looked to be incorrect as a result of the creaky mode being incorrectly identified according to the original procedure, the identification of a new creaky mode led to a new antimode being identified. The exception to this in both cases was Shetland OF Betty, whose creaky mode and antimode remained the same. However, as described in Section 7.3.5.3, Betty's speech included both cases where PPA-identified creak was not identified because it occurred above the antimode, and cases where she used non-creaky voice below her antimode, meaning that a lower antimode may have caused more issues. The new procedure did not identify new antimodes for the two speakers whose creaky modes had been identified correctly according to the original procedure.

Modes and antimodes for speakers where these changed are given in Table 7.5.

Table 7.5: New modes and antimodes for speakers where these changed using the new automated procedure. Changes are presented in **bold.**

Area	Pseudonym	Antimode	Non-creaky mode	Creaky mode
Glasgow	Dale	67	96	**51**
Glasgow	Scott	**69**	93	**49**
Lothian	Tina	110	163	**82**
Shetland	Kellie	**130**	206	**106**
Shetland	Stephanie	138	167	**85**
Shetland	Alexander	79	117	**61**
Shetland	Graham	**79**	113	**58**
Shetland	Hugh	65	94	**32**

7.3.6.2 Changes to disagreements for these speakers

With the new, lowered antimode thresholds for these speakers, 27 pulses from 14 different stretches which were originally coded as creaky using the f0-based antimode method were no longer coded as creaky. At the level of the pulse, this led to the overall level of agreement lowering, with 17 of these pulses changing from 'Agree' to 'Disagree', and 10 pulses changing from 'Disagree' to 'Agree'.

Many pulses that were re-coded under the new procedure occurred within stretches that already contained a disagreement between the two schemes. This meant that despite the fact that agreement changed at the level of the GCI for 27 pulses, agreement at the level of the stretch only changed in four cases. One stretch changed from having complete agreement to containing a disagreement, while three stretches changed from containing some disagreements to agreeing completely.

7.4 Discussion of F0-based identification of creak

In this chapter, I investigated how PPA coding of creak related to f0-based identification of creak, where creak was identified where f0 fell below a speaker-specific threshold, the antimode, identified using an automated procedure following Dorreen (2017), Dallaston & Docherty (2019) and Szakay & Torgersen (2019).

F0-based identification of creak originates in Dorreen's (2017) research on language-specific f0 variation in bilingual speakers, which found REAPER to be more accurate than Praat's method across wider f0-ranges without needing to be calibrated to a specific f0 range. This gives it the advantage of consistency across studies and allows it to be used to identify f0 modes for creaky and non-creaky speech without separate calibration. Dallaston & Docherty (2019) then implemented this method as a coarse-grained method of identifying creak, and considered its efficacy in 42 Australian English speakers aged 18-50. More recently, White et al. (2022) have compared the performance of this method to other automated approaches of identifying creak.

Following this method, I found that in 87% of voiced stretches, F0-based identification of creak agreed with PPA coding over the entire stretch. In half of the remaining 13% of stretches, the two system agreed on the creak status of at least 75% of the GCIs. I investigated the reasons behind disagreements between the two systems, and found that most often differences between the two systems arose because of time resolution differences between the two systems, arising from their different units of analysis. However, several other reasons also accounted for half of the disagreements and some of them present issues with the f0-based coding system using REAPER. Here, I compare my findings to those of Dallaston & Docherty (2019) and White et al. (2022), and consider disagreements between the system in more detail. I also consider what can be done to mitigate issues with this method in the larger corpus study.

To recap, Dallaston & Docherty (2019) found that f0-based coding agreed with manual coding in 98% of non-creak and 81% of creak. These figures are derived from analysis where both manual and f0-based coding operated using GCI units, so I compare my analysis at this level. With this in mind, I find the same percentage of f0-based non-creak was coded as non-creaky in PPA, at 98% (62,615 cases of positive agreement out of 63,963 total non-creaky GCIs according to the f0-based method). However, my analysis finds a lower percentage of f0-based creak was identified manually, 70% compared to Dallaston & Docherty's 81%.

White et al. (2022) compared based on creak coded by manual annotators in 10ms frames to creak coded by the f0-based method. Of their 386,576 frames, less than 1% were false positives (=Antimode-only), 3% represented false negatives (=manual annotation only), 3% represented true positives (=positive agreement), and 93% rep-

resented true negatives (=negative agreement). My own analysis works at the level of the GCI and the voiced stretch, rather than in 10 ms frames, meaning that only the analysis transformed into ms can be compared to White et al. (2022). By comparison, I found similar levels of antimode-only creak (<1%), but more negative agreement (97%), less positive agreement (1%), and less manual-only coded creak (2%). This mostly suggests that Scots speakers are less creaky than Australian English speakers, and that this persists in both manual and f0-based coding.

White et al. (2022) also present their findings in terms of $F1$ score to account for the fact that creak occurs less often than non-creak. This approach weights positive agreement (or true positives, in White et al.'s framing) more highly than other types of agreement, and does not focus on the large amount of negative agreement that occurs. White et al. (2022) modelled $F1$ score to compare it across several different automatic methods of identifying creak and across male and female speakers. Their model predicted an $F1$ score of 0.69 for female speakers, and 0.42 for male speakers.

Participants in White et al.'s (2022) study were aged 18-34, so are comparable to the younger speaker group in this study. *F1* for younger female and male speakers in this study was higher than in White et al.'s (2022) study, at 0.77 and 0.53 respectively. However, performance was considerably worse among the older speakers in this thesis, with older female speakers showing an $F1$ of 0.43 and older male speakers showing a very poor $F1$ at 0.17.

Some potential reasons for the lower performance in male speakers and older speakers are found in comments in previous research. Dallaston & Docherty (2019) remark that most cases of creak identified by the f0-based method but not by manual coders could be attributed to either creak occurring over a short time-span so that it was not picked up by manual coders, or incorrect GCI detection. Furthermore, they attribute cases of manually-annotated creak missed by the f0-based method to the failure of REAPER to identify GCIs of very low F0 and very low intensity, and cases characterised by irregular rather than low f0. Furthermore, White et al. (2022) remark that as male speakers tend to use lower f0 in their non-creaky f0 range, they are likely to have less robust antimode separation. They also speculate that f0-based method might be picking up to creak that human annotators missed in male speech.

In the present study, hand-checking of disagreements between the two coding systems allowed these possibilities to be investigated more closely. The largest category of disagreements, which included differences in time resolution supports Dallaston & Docherty's (2019) remark short instances of creak are sometimes missed in manual coding. These disagreements reflected the different units of analysis of each method: In f0-based coding, analysis took place at the level of the GCI, while in PPA, it took place at the level of the voiced stretch. Furthermore, they reflected a difference in the

criteria used to code a stretch as creaky — f0 alone in the f0-based method, compared to PPA's auditory approach, in which I paid attention to the overall impression in the voiced stretch on the basis of not just pitch, but auditory creaky quality. These cases often tended to be cases of only a few pulses of segmental creak, often occurring at the beginning of the stretch in the form of a glottal onset, or at the end of a stretch due to a following glottal or instance of pre-glottalization. In PPA, these cases were excluded from the stretch if more than two pulses occurred but the stretch of creakiness was too short to be counted as its own stretch. In practice, this meant that many cases of segmental creak with only one or two pulses were included in PPA voiced stretches that were not coded as creaky. Relevant to the larger corpus study presented in Chapter 9, the fact that this is the most frequent type of disagreement supports the use of this method with more data, with the caveat that the method operates at a different time span than human coding of creak.

The second largest category of disagreements between the two systems arose because of f0 tracking errors in REAPER, supporting Dallaston & Docherty's (2019) observation of cases of incorrect GCI detection in cases of low and irregular f0, and low intensity. Interestingly, numerical analysis of disagreement types reveals that f0 tracking errors rarely lead to creak being missed by the f0-based method, and more often lead to creak being coded in non-creaky speech. This suggests that REAPER is possibly more likely to produce pitch-halving errors in non-creaky speech than pitch-doubling errors in creaky speech. This is an important limitation in the present research as the limited flexibility of REAPER leaves limited options for decreasing the number of f0-tracking errors when the method is rolled out to the wider corpus.

A novel finding is the poorer performance of REAPER in non-modal voicing, such as highly whispery voice. This is a concern in the present research, as the aim of using f0-based identification of creak is to separate creak from non-creak in order to analyse the two separately. If highly amodal, non-creaky speech causes f0 tracking errors (and false identification of creak), this is a significant limitation in this study.

f0 tracking errors disproportionately affect older male speakers. This suggests that REAPER encountering issues identifying GCIs in older male speakers is the main reason underlying the lower $F1$ score in older male speakers. This will be a significant limitation in the larger corpus study presented in Chapter 9, as it limits the extent to which automatically identified creak is comparable between older male speakers and other age/gender groups.

However, it is worth noting that f0 tracking errors do not always result in incorrect creak identification using the antimode method; rather, I found multiple cases where creak caused issues with f0 tracking, lowering f0 below the antimode, and causing creak to be identified.

The next most common type of disagreement involved antimode errors (where creak occurred above the f0 antimode, or where non-creak occurred below the antimode). These were far more likely to miss creak coded in PPA than to code non-creaky speech as creaky. In the case of one speaker, Gary (OM Lothian), this appeared to be caused by an interaction between inaccurate f0 tracking and the antimode method, which led the antimode to be artificially lowered. Furthermore, for some speakers, speaker-specific f0 antimodes either cannot be identified or do not consistently distinguish between creaky and non-creaky voicing. This was particularly evident in the case of Margaret (OF Lothian), where an f0 antimode between creaky and non-creaky speech could not be identified, predominantly due to smooth f0 transitions between non-creaky and creaky voicing, as well as in the case of Betty (OF Shetland), whose non-creaky f0 range and creaky f0 range overlapped despite an identifiable antimode. Potentially, using more data from each speaker could help to resolve these issues. Although this would never be able to produce 100% separation between creaky and non-creaky speech in cases where creaky and non-creaky f0 ranges overlap, it may be that with more data an antimode will appear that will produce some separation between the two types of voice, and that separation will increase with more data. Another avenue would be to look for local f0 antimodes: While there was no singular value that would consistently separate creak for Betty, creak does still appear to have a lower f0 than non-creaky voice from within the same utterance, so breaking the recording down into utterances could reveal a local f0 antimode that separates them on a smaller scale. Doing this would likely restrict the extremities of a speaker's f0 ranges so that the lower end of a speaker's non-creaky range would be less likely to overlap with the top of their creaky range.

Antimode errors disproportionately affect older female speakers, and occur slightly more in younger male speakers than younger female speakers. This suggests that White et al. (2022) speculation that this method performs worse in male speakers due to more unimodal f0 distributions may have been the case in their data, as among speakers of a comparable age group male speakers are slightly more affected by this issue. However, when this method is expanded to speakers over 65, it is female speakers who show more issues with automatically identified creak relating to unimodal f0 distributions. This is supported by Margaret's (OF Lothian) unimodal f0 distribution, as well as by near-unimodal f0 distributions in other older female speakers such as Valerie (OF Glasgow). I aimed to mitigate this by implementing the new version of the automated procedure, which tests potential creaky modes in order of height rather than distance from the non-creaky mode, but this did not result in an antimode being identified for Margaret and only changed the evaluation of a small amount of pulses.

In future research using the f0-based method, acoustic measurement of automatically coded creaky speech using measures such as H1*—H2* and CPP might be useful to discern whether creak has been broadly identified correctly without conducting ad-

ditional auditory-perceptual investigation. This might also help to reveal what types of creak are being used, which would allow investigation into whether different types of creak may be used for different social and linguistic purposes.

Though outside of the scope of the present study, in future work, it could be worth investigating the use of the same methodology with alternative f0 trackers. While REAPER is known for being reliable at low and irregular fundamental frequencies, other reputable f0 trackers such as STRAIGHT have not yet been used for estimating creak in this method but could have potential to decrease the number of tracking errors that result in creak being identified incorrectly.

A less common issue in this corpus which could cause more issues elsewhere was cases where speakers used creak that was characterised by multiple pulsing rather than low f0. One potential method for differentiating this in future research could be to use the local difference between consecutive GCIs to identify creak: If multiple pulsing characterises creak and this is accurately tracked by REAPER, then creak should be identifiable by measuring f0 variation between pairs of consecutive cycles. However, this would require an f0 tracker that could reliably identify all GCIs in cases of multiple pulsing, which is not the case with REAPER.

White et al. (2022) present a method that unifies the f0-based approach with the resonator-based method presented by Kane, Drugman & Gobl (2013), so that creak is identified where either method would have coded it. They find this method performs better at detecting creak than either method in isolation. While this method was still under development when the analysis in this thesis was being conducted, it may be a useful direction for future research.

Overall, performance of the f0-based method in speakers with Scottish accents appears roughly comparable to the performance identified in previous research. The fact that the majority of cases where PPA coding and f0-based coding disagreed can be attributed to time resolution differences is promising for the present study, where this method will be rolled out to a larger dataset in Chapter 9. However, f0 tracking errors and antimode errors that arise from unimodal f0 distributions are significant issues for the current research, especially their disproportionate effect on the speech of older male and older female speaker respectively. In the larger corpus, I hope that the use of more data per speaker and the new automated procedure will help to mitigate antimode problems, but no strategy has been identified to aid in issues that arise from f0 tracking problems. Instead, I hope that combining this method with analysis of non-creaky tokens using VoiceSauce should also give an idea of whether there is creak being used in data that is not coded as creaky using the f0-based method. This will be particularly relevant for speakers like Margaret, whose f0 antimode could not be identified, but who still use creak. This will allow results to potentially be interpreted

as creak in the face of high noise and low spectral tilt, despite lack of automated f0-based identification.

Chapter 8

Connecting PPA and acoustic analysis of non-creaky voice quality

8.1 Introduction

In Chapter 6, I presented an auditory-perceptual analysis of voice quality in 24 speakers from SCOSYA using PPA. I found that voice quality in Scottish accents was characterised by whispery and tense whispery voice. Glasgow was characterised by tense whispery voice, while Lothian used more near-modal qualities. Younger speakers (18-25) used more modal voice and less harsh voice than older speakers (65+). I also found gender variation, with male speakers using more breathy voice, harsh voice and higher scalar degrees of breathy and whispery qualities, and female speakers using more whispery, tense and tense whispery voice. In Chapter 7, I considered how automatic coding of creak related to auditorily coded creak in PPA, enabling this analysis to be scaled up in Chapter 9 to consider how creak patterns according to social and linguistic factors in Scottish voice quality. In this chapter, I aim to connect PPA labels for other qualities to acoustic measures of voice quality, enabling the acoustic analysis of Scottish voice quality conducted in Chapter 9 with more data and more speakers to be interpreted with reference to descriptive phonetic terminology for voice quality.

8.1.1 Interpreting acoustic measurements in terms of auditory-perceptual descriptions of voice quality

In Section 4.2, I gave an overview of different acoustic measures. Here, I return to spectral tilt and noise measurements and explore how I expect them to relate to the auditory-perceptual labels of PPA.

8.1.1.1 Harmonic source spectral shape

I took three measures of harmonic source spectral shape, which each related to a different part of the spectrum: H1*–H2*, H2*–H4* and H4*–2kHz*. In forming my predictions, I make reference to Zhang's (2016a) three-dimensional computational model of voice production, which aimed to establish how vocal fold stiffness, medial surface thickness, glottal aperture, and subglottal pressure affect resulting acoustic measures. Note that here I omit the asterisk signifying formant correction, as I refer to the theoretical effect of phonation on the harmonics, rather than a measured and formant-corrected effect.

H1–H2 has been widely used in voice research. Zhang (2016a) links glottal configurations with higher airflow, a longer open phase, and less contact between the vocal folds to H1 being more dominant in the low-frequency portion of the spectrum. This leads to a greater difference between H1 and H2, and ultimately larger values

for H1–H2. These configurations are typical of breathy voice, produced with minimal adductive tension, medial compression and longitudinal tension (Laver 1980) and no epilaryngeal constriction (Esling et al. 2019). Configurations used in the production of other phonation types may also increase H1–H2. The glottal configuration in whispery voice resembles that of breathy voice, with 'progressive tightening of the laryngeal constrictor generat[ing] increasing degrees of whispery voicing' (Esling et al. 2019: 58), which in turn constrains vocal fold vibration to the anterior portion of the vocal folds and increases medial compression. High airflow, a longer open phase and lower vocal fold contact are still typical of whispery voice, in turn enhancing the first harmonic.

Thinner vocal folds increase H1–H2 (Zhang 2016a), so H1–H2 should also be higher in falsetto and 'laryngeal constriction at high pitch' (Esling et al. 2019), two phonation types where the vocal fold thickness is reduced. Zhang (2016a) finds that increasing medial surface thickness increases vocal fold contact, reduces airflow through the glottis, and shortens the open phase. This in turn excites harmonics above H1, leading to a peak in the low-frequency portion of the spectrum above H1 and a reduction in the attenuation in amplitude as frequency increases. H1–H2 should therefore be reduced in phonation types with higher vocal fold contact, such as modal voice and tense voice, and particularly so for creaky voice, produced with thicker vocal folds.

However, a shorter open phase leading to a peak in the low-frequency portion of the spectrum may also mean that H4 is enhanced, in turn reducing H2–H4. Though less widely applied than H1–H2, H2–H4 has also been found to differentiate between breathy and modal vowels in several languages, such as Gujurati (ud Dowla Khan 2012), as well as Chong, Fuzhou, Mon, and San Lucas Quiaviní (Esposito 2010).

Above H4, spectral slope also relates to the abruptness of vocal fold closure. Zhang (2016a) illustrates how increased angle at the glottis means that posterior portion of the vocal folds has to travel further to the midline, resulting in glottal closure taking longer and being less abrupt. This reduces the excitation of higher-order harmonics, increasing spectral slope. This could mean that posterior glottal gap typical of whispery voice may increase spectral slope in the H4–2kHz region. By contrast, tense and creaky voice may be characterized by a less steep spectral slope in this region, due to the more abrupt cut-off between the open and closed phases of the glottal cycle.

The present study uses H1*-H2*, H2*-H4* and H4*-2kHz* to examine differences in spectral shape between whispery, breathy, modal, tense, and tense whispery voice. Figure 8.1 shows the expected spectral shape of each of these voice qualities. I expect whispery voice, the reference level in the model comparing phonation types, to have a moderately steep slope in the H1–H2 and H2–H4 regions reflecting the high airflow, long open phase and incomplete glottal closure of whispery voice, and a steep slope in the H4–2kHz region, reflecting the glottal gap in the posterior vocal folds. By comparison,

I expect breathy voice to have a steeper slope in the H1–H2 and H2–H4 regions than whispery voice. In the H4–2kHz region, breathy voice may also exhibit steep spectral slope. An alternative possibility, depicted here, is that H4–2kHz may be less steep for breathy voice than for whispery voice, which would suggest a less 'Y-shaped' glottal configuration for breathy voice, with a similar degree of incomplete closure posteriorly and anteriorly.

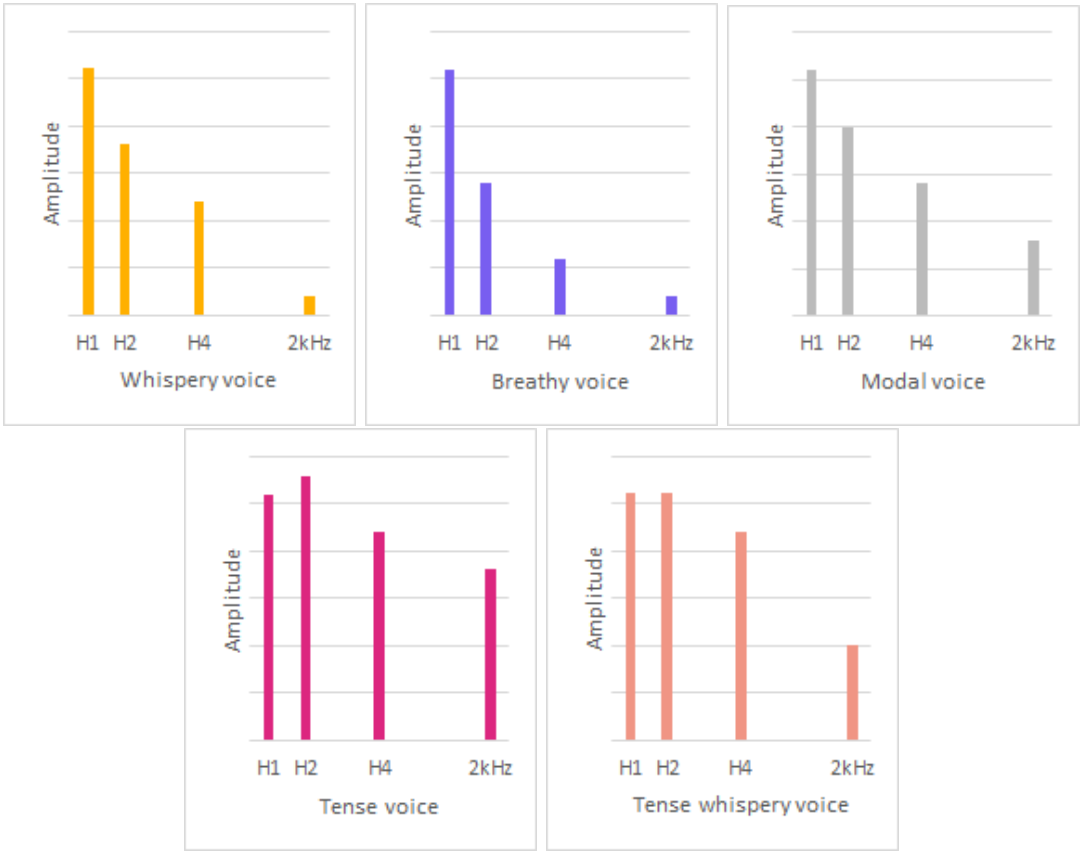


Figure 8.1: Schematic spectra illustrating the expected spectral shape between H1 and 2kHz for the voice qualities considered in the between-category analysis

In comparison to whispery voice, modal voice should have more vocal fold contact, a shorter open phase and theoretically no glottal gap. This configuration would cause the spectral slope to fall off at -12 dB/octave (Flanagan 1957), decreasing spectral slope relative to whispery voice.

Tense voice is characterised by a shorter open phase than whispery voice, and abrupt glottal closure. I expect this to cause a flatter spectral slope than in other phonation types considered here. The shorter open phase may cause a peak in the H2 or H4 region (depicted in the H2 region here), causing negative values for H1–H2 or H2–H4, while H4–2kHz may be flatter because of more abrupt vocal fold closure.

The configuration underlying tense whispery voice and the resulting spectral shape is less clear. Laver (1980: 146) states that the addition of whispery voice to tense voice results in a configuration where 'the cartilaginous glottis is kept open against adductive tension only by vigorous antagonistic action'. The cartilaginous glottis that

Laver refers to is the space in between the arytenoid cartilages, suggesting that tense whispery voice would maintain the posterior glottal opening of whispery voice. Despite this, he states that the laryngeal tension in tense whispery voice continues to be very audible; it is uncertain how this tension may manifest acoustically. One possibility is that it may lie between whispery and tense voice in terms of spectral tilt. Alternatively, it could show some characteristic features of one component and some of the other: For example, it may be produced with a shorter open phase than whispery voice, leading to a flat slope between H1–H2 and H2–H4 and an auditory tension component, but a posterior glottal gap, creating a steep slope between H4–2kHz, depicted here.

I also expect differences between increasing scalar degrees of breathy and whispery voice. For breathy voice, I expect that an increasingly breathy percept will be related to increased spectral slope. However, for whispery voice, I expect that an increasingly whispery percept will be more related to noise than spectral slope.

8.1.1.2 Noise

Initial intention was to model noise using multiple measures which consider different parts of the frequency range. However, due to issues with violation of model assumption using multiple HNR measures, only CPP was used in the final analysis. CPP is described in Section 4.2.3, but to recap, it compares the amount of aperiodic noise and harmonic energy in the cepstral domain. Higher values indicate more harmonic energy in the signal, while lower values indicate more aperiodic noise in the signal. Broadly, this means that more modal-like phonation has higher CPP, while less-modal qualities will have lower values.

I expected whispery voice to involve a high level of noise (=high CPP), resulting from incomplete vocal fold closure and epilaryngeal constriction. Tense whispery voice and breathy voice will likely include a degree of noise, but less noise than whispery voice. By comparison, modal and tense voice are likely to minimal noise (=low CPP).

8.2 Methods

Parameters were estimated using VoiceSauce (Shue et al. 2011) with a 16 kHz sampling rate and were taken over time for each voiced stretch, then averaged within each stretch. The following measurements were taken:

- Snack F0 (Sjölander 2004)
- STRAIGHT F0 (XSX) (Kawahara et al. 2008)
- F1, F2, F3, F4 from Snack, using STRAIGHT f0 estimate
- B1, B2, B3, B4 from Snack – measured, not estimated from formula
- H1*–H2*, H2*–H4*, H4*–2kHz* (corrected using formant frequencies and bandwidths calculated as above)
- Cepstral Peak Prominence (CPP)
- HNR05
- HNR15
- HNR25
- HNR35

8.2.1 Trimming of data points and outliers

Stretches that were coded as 2 degrees or higher for creaky voice in PPA were excluded. Tense voice (Stretches rated as 1 for creaky voice), however, is considered here — Referred to from here on as 'tense voice'. After excluding creak, 1961 stretches remained.

Because reliable estimation of harmonic measurements requires accurate f0 measurements, and because difficulty tracking f0 may indicate that other parameters may be difficult to estimate, f0 estimates were trimmed by taking the difference between STRAIGHT (Kawahara et al. 2008) f0 estimates and the REAPER (Talkin 2015) f0 estimate, averaged over the course of the stretch, and discarding data points where the difference between the two exceeded 5 Hz. As shown in Figure 8.2, 89.8% of data (1761 tokens) fell within this range, represented by the dotted pink lines in the plot.

Furthermore, because formant-correction of harmonic measurements requires accurate formant measures, I also trimmed on the basis of formant measurements. The first quartile (Q1) and the third quartile (Q3) contain the middle 50 % of the data

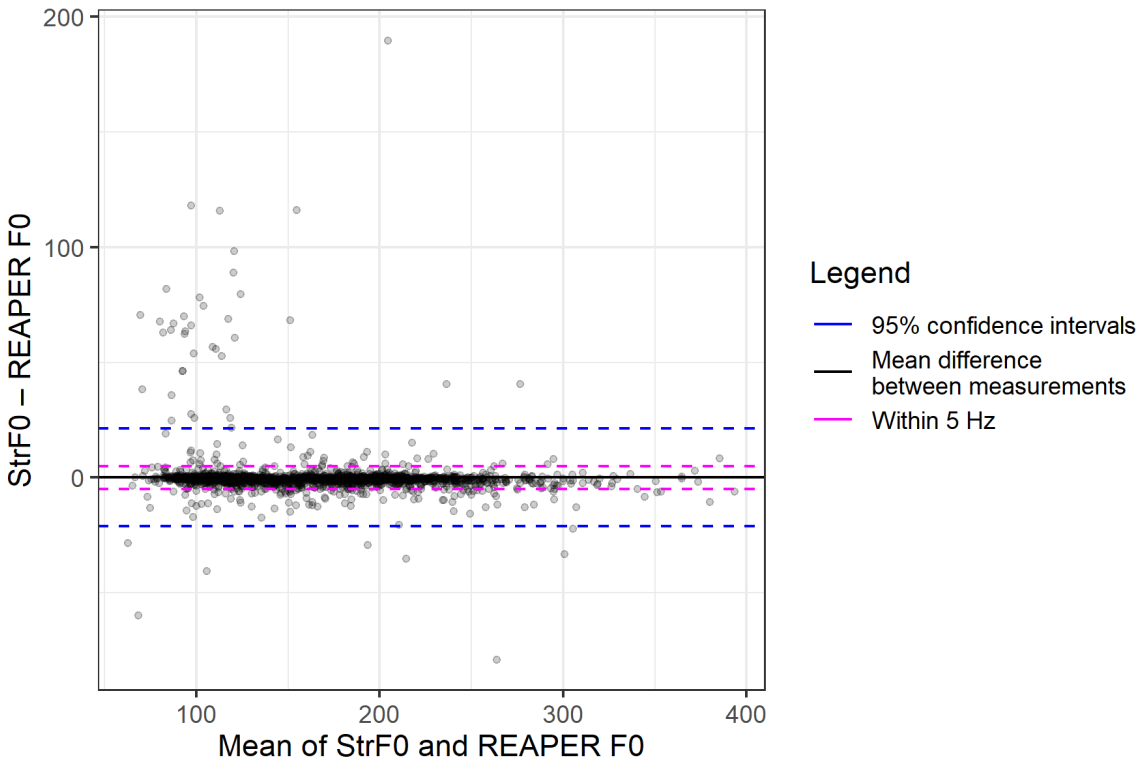


Figure 8.2: Bland-Altman plot showing the mean of STRAIGHT and REAPER F0 measurements and the difference between the two measures.

surrounding the median, and the range between these two quartiles is known as the inter-quartile range (IQR). I defined outliers here as lying beyond 1.5*IQR below Q1 or above Q3, and trimmed values for F1, F2 and F3 which fell outside of these ranges. Following trimming on the basis of formant values, 1704 voiced stretches remained for analysis.

8.2.2 Analysis of categorical coding of voice quality

8.2.2.1 Dependent variable: PPA-coded voice quality

In PPA, each voiced stretch had been coded for the presence or absence of each of the following voice qualities:

- Whispery voice
- Tense voice
- Breathy voice
- Harsh voice
- Modal voice

- Falsetto
- Whisper

Voiced stretches could be coded as exclusively voice quality, or could be coded for the presence of a combination of different voice qualities (e.g. tense whispery voice). This analysis the probability of a voiced stretch being coded as a certain voice quality, as a function of a set of acoustic measures taken from that voiced stretch.

After data were trimmed as described in Section 8.2.1, this resulted in the distribution of voice qualities across all voiced stretches as shown in Table 8.1.

Voice quality category	Count
Whispery voice	550
Tense whispery voice	511
Breathy voice	335
Tense voice	162
Modal voice	118
Whispery harsh voice	18
Harsh voice	5
Breathy + creaky (lax creak)	2
Whispery harsh creaky voice	2
Harsh creaky voice	1
Whisper	0
Total	1704

Table 8.1: Count of each voice quality category

I analysed this using multinomial logistic regression, and an extension of binary logistic regression that allows an outcome variable with multiple categories to be predicted as a function of one or more independent variables. The model was run using the `multinom()` function from the package `nnet` (Venables & Ripley 2002).

Because multinomial logits assume that a data point cannot belong to more than one category, categories that represented overlap between two categories (e.g. tense whispery voice) were considered to be a separate category and treated as mutually exclusive (e.g. a tense whispery token would only be included in the tense whispery voice category, and not in the whispery or tense category) Furthermore, because the voiced stretches were not evenly distributed between each voice quality category, $n = 50$ was taken as a cut-off point, below which a voice quality category was not included in the model. Because of this, only the following voice quality categories were included in the model:

- Whispery voice
- Tense whispery voice

- Breathy voice
- Tense voice
- Modal voice

8.2.2.1.1 Baseline category Multinomial logistic regression can be conceptualised as a series of binary logistic regressions that are run to compare levels of an outcome variable to a baseline category of that variable, resulting in a series of regressions equivalent to one fewer than the number of levels in the outcome variable (Wiley & Wiley 2019: 142). As whispery voice contained the highest number of tokens, whispery voice was taken as the baseline.

8.2.2.2 Independent variables: Acoustic measures

The intention of this model was to predict the probability of a voiced stretch being coded as a certain voice quality, as a function of a set of acoustic measures taken from that voiced stretch. I had therefore initially planned to include the following independent variables in the model:

- H1*–H2*
- H2*–H4*
- H4*–2kHz*
- HNR05
- HNR15
- HNR25
- HNR35
- CPP

Values for each of the independent variables were scaled and centred around the mean. The data was trimmed further on the basis of the values of the independent variables, so that data points below Q1 - 1.5*IQR or above Q3 + 1.5*IQR were discarded. This left a final total of 1,669 tokens for analysis.

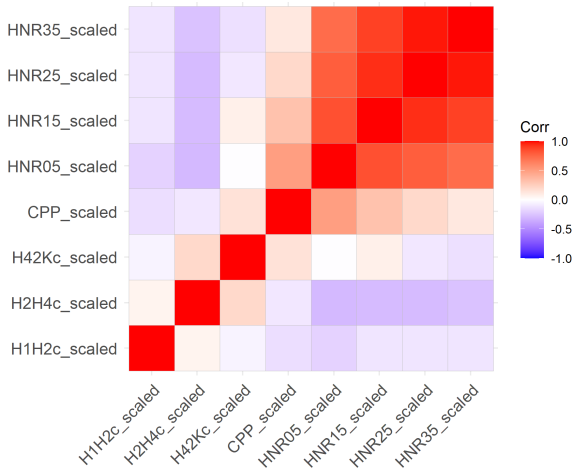


Figure 8.3: Correlation plot showing the degree of correlation between the different acoustic measures in the analysis of categorical coding of voice quality

8.2.2.3 Multicollinearity

However, one assumption of the multinomial logit is that there is no multicollinearity between the independent variables. Following Belsley, Kuh & Welsch (1980), I checked for multicollinearity by calculating the condition number for the proposed independent variables and found medium multicollinearity ($c = 20$). Because different noise measures (HNR05, HNR15, HNR25, HNR35, CPP) overlap in the parts of the spectrum that noise is calculated for, I suspected that it was noise measures that were responsible for this. I confirmed this by looking at the correlations between different measures: As shown in Figure 8.3, the HNR measures (HNR05, HNR15, HNR25, HNR35) were highly correlated with each other, having a correlation coefficient of 0.76 or above, and CPP also showed a degree of correlation with some HNR measures. I therefore constructed models containing the spectral tilt measures and each noise measure and selected the final model by comparing the AIC for each model, and found that the model that contained spectral tilt measures and CPP (but no other noise measures) had the lowest AIC. I then calculated the condition number for the model containing spectral tilt measures and CPP and found there was no longer multicollinearity in the data ($c = 1$).

8.2.2.4 Independence of Irrelevant Alternatives (IIA) Assumption

Another assumption of multinomial logit models is the Independence of Irrelevant Alternatives (IIA) Assumption. This assumes that the chance of selecting a particular outcome in the model is irrelevant of the presence or absence of other options. For example, if I were to model the chance of people voting for Labour or SNP in a

constituency in a general election, it would be reasonable to assume that the presence or absence of a Conservative candidate on the ballot might affect voters' choice to select Labour or SNP, as voters might choose to vote tactically to block a Conservative candidate: This would cause the assumption to be violated. On the other hand, the presence of a joke candidate on the ballot would be unlikely to affect voters' choice to select Labour or SNP: In this case, the alternative option would be irrelevant to voters, and the assumption would be met.

A more relevant example is that of having had an additional 'not sure' category in PPA. In this case, I might have placed edge cases into the 'not sure' category, then only revisited more certain or extreme cases when rating future cases, which may have affected the perceptual boundary between other categories. This would then mean that the existence of a 'not sure' was not independent of my choice to place a stretch into one category over another.

In practice, I wanted to test whether having the option to rate a voice as tense, tense whispery voice, modal, or breathy had affected my ratings for other categories. I tested the IIA assumption using the Hausman-McFadden test (Hausman & McFadden 1984) implemented in R in the `mlogit` package (Croissant 2020), comparing versions of the model which included and excluded different levels of the outcome variable. For each iteration of the Hausman-McFadden test, $p > 0.05$, meaning that there was no reason to reject the null hypothesis and suggesting that the IIA assumption was met.

8.2.2.5 Model selection

The full model was then stepped down automatically using the `step()` function in R Core Team (2020) which adds and removes predictors and compares each model using Akaike's information criterion (AIC) in a Stepwise Algorithm to choose an optimal model. However, removing predictors did not improve the model, so the model was selected as follows:

1. `Voice quality ~ H1*--H2* + H2*--H4* + H4*--2kHz* + CPP`

8.2.3 Analysis of the scalar degree of breathy voice

In PPA, tokens that were coded as breathy voice were given a scalar degree between 1 and 5, where 1 was equivalent to lax phonation and 5 was maximally breathy.

Only voiced stretches that had been coded as exclusively breathy were considered. After trimming as described in Section 9.1.1, 335 voiced stretches met this criteria. There were no outliers (defined as data points below Q1 - 1.5*IQR or above Q3 +

1.5*IQR for any of the possible independent variables) and so no further trimming was conducted.

For each of these stretches, breathy voice had been rated as a scalar degree between 1-5 as shown in Table 8.2.

Scalar degree	Count
1	99
2	77
3	111
4	35
5	13
Total	335

Table 8.2: Count of each scalar degree of breathy voice

Because of the low number of tokens rated as scalar degree 4 and scalar degree 5, these categories were grouped together into scalar degree 4.

Because the within-category ratings were completed on an ordinal scale, I decided that an ordered logistic regression model was the most appropriate, as ordered logits are used to model dependent variables with multiple categories where these categories have a natural ordering as a function of one or more independent variables. However, ordered logistic regression also assumes that the degree of change between each level of the dependent variable is proportional. Described in further detail in Appendix B.2, this was not the case for all independent variables. I therefore conducted the analysis with multinomial logistic regression.

As for the analysis of categorical phonation types, I initially planned to include the full set of acoustic measures as independent variables in the model, but encountered some issues of multicollinearity between noise measures. I checked for multicollinearity by calculating condition number for the proposed independent variables and found medium multicollinearity ($c = 19$). I then inspected the correlation coefficients between the different acoustic measures to see if the issue appeared to be caused by multicollinearity between different noise measures. As shown in Figure 8.4, the different HNR measures were highly correlated with each other, with the correlation coefficients between each pair of HNR05, HNR15, HNR25 and HNR35 being at least 0.74. CPP also showed a degree of correlation with HNR measures, with the correlation coefficient between CPP and HNR05 being 0.51.

I then constructed models containing the spectral tilt measures and each noise measure in turn as compared the AIC for each model, and proceeded with the model containing only CPP as a noise measure as it had the lowest AIC. I then recalculated the condition number and found that multicollinearity was no longer an issue ($c = 2$).

As in Section 8.2.2.4, I checked that the data met the IIA assumption by using the

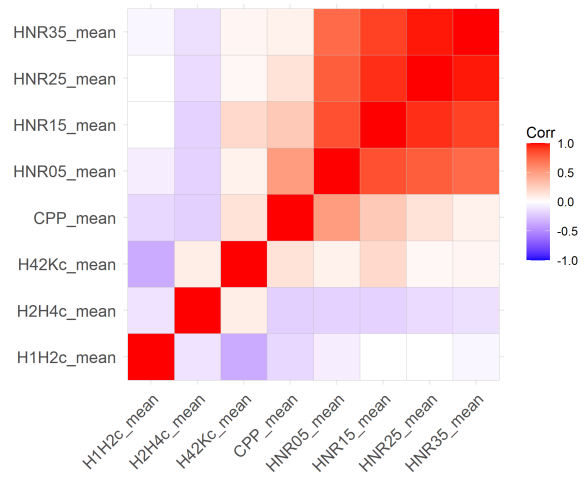


Figure 8.4: Correlation plot showing the degree of correlation between the different acoustic measures in the analysis of breathy voice

Hausman-McFadden test. For each iteration of the test, p >0.05, meaning that the assumption was met.

The full model containing H1*–H2*, H2*–H4*, H4*–2kHz* and CPP was then stepped down automatically using the `step()` function in R Core Team (2020). Removing H4*–2kHz* from the model improved the AIC. Accordingly, the final multinomial logit modelling scalar degrees of breathy voice as a function of acoustic measures was:

1. `Breathy voice ~ H1*-H2* + H2*-H4* + CPP`

8.2.4 Analysis of scalar degrees of whispery voice

This analysis considered how the scalar degree of whispery voice varied as a function of a set of acoustic measures taken from that voiced stretch. Only voiced stretches that had been coded as exclusively whispery were considered. After trimming as described in Section 9.1.1, 550 voiced stretches met this criteria. I then trimmed outliers (below Q1 - 1.5*IQR or above Q3 + 1.5*IQR) for each of possible independent variables, which resulted in trimming one token with H1*–H2* below this range. For each of these stretches, whispery voice had been rated as a scalar degree between 1-5 as shown in Table 8.3. As before, scalar degree 4 and 5 were collapsed into a single category.

As before, I had initially planned to include the full set of acoustic measures as independent variables in the model. Following the same procedure as for the multinomial logit for breathy voice, I tested for potentially harmful multicollinearity. Again,

Scalar degree	Count
1	89
2	168
3	215
4	62
5	15
Total	549

Table 8.3: Count of each scalar degree of whispery voice

medium multicollinearity was an issue in the full model ($c = 19$). I proceeded with a model containing H1*–H2*, H2*–H4*, H4*–2kHz* and HNR05. I again calculated the condition number to check that there was no potentially harmful multicollinearity in the data.This resulted in the same set of measures being selected: H1*–H2*, H2*–H4*, H4*–2kHz* and CPP.

As in Section 8.2.2.4, I checked that the data met the IIA assumption by using the Hausman-McFadden test. For each iteration of the test, p >0.05, meaning that the assumption was met.

The full model containing H1*–H2*, H2*–H4*, H4*–2kHz* and CPP was then stepped down automatically using the `step()` function in R Core Team (2020). Removing H1*–H2*, H2*–H4* and H4*–2kHz* from the model improved the AIC. Accordingly, the final multinomial logit modelling scalar degrees of whispery voice was:

1. `Whispery voice` $\sim$ `CPP`

8.3 Results

In this section, I present the results of three multinomial logistic regression models that consider how PPA coded voice quality varies according to a range of acoustic measures. Multinomial logistic regression models the effect of one or more independent variables on a dependent variable with more than two categories. Unlike linear regression and binary logistic regression, multinomial logistic regression is rarely used within sociolinguistics and phonetics, so many readers of this thesis may not be familiar with how to interpret the results of a model of this kind. To guide interpretation of the results of models, I present an worked example of how to interpret multinomial logits using fake data in Appendix B. Here, I present the actual results of the multinomial models.

8.3.1 Analysis of categorical coding of phonation type

Table 8.4 presents the results of the multinomial logit model predicting log(odds) of a voiced stretch being rated as each phonation type, given H1*–H2*, H2*–H4*, H4*–2kHz*, and CPP. All independent variables have been mean-centred and scaled. The first main row of the table shows the log(odds) of a stretch being rated as breathy voice, modal voice, tense voice, or tense whispery voice, rather than whispery voice, followed by the standard error (SE) on the next row in bracket, and the t-value and asterisks to indicate significance on the third row.

8.3.1.1 H1*–H2* and PPA-coded phonation type

To begin, we will consider the effect of H1*–H2* on how voice quality is categorised. An increase of 1 SD in H1*–H2* increases the log(odds) of a voiced stretch being categorized as breathy rather than whispery by 0.280, while it decreases the log(odds) of a voiced stretch being categorised as tense rather than whispery by 0.085, and has no significant effect on the log(odds) of a stretch being categorised as modal or tense whispery.

The effect of H1*–H2* on how voice quality is categorised can be seen in Table 8.5 which shows the predicted probability of a voiced stretch being rated as each phonation type at a range of values for H1*–H2*, where all other independent variables in the model are held constant at their means. At 2 SD below the mean, the model predicts that the probability of a voiced stretch being categorised as breathy is 10%. This increases to 18% at the mean, and to 29% at 2 SD above the mean. This can be seen in Figure 8.5 by the increasing width of the purple panel as H1*–H2* increases, representing a higher predicted probability of breathy voice as H1*–H2* increases.

Table 8.4: Summary of multinomial logit model predicting phonation type as a function of H1*–H2*, H2*–H4*, H4*–2kHz*, and CPP. In each cell the estimate is presented first in log-odds, followed by the standard error (SE) in brackets, then the t-value and any asterisks indicting statistical significance.

	Dependent variable:			
	breathy	modal	tense	tense whispery
	(1)	(2)	(3)	(4)
Constant	−0.719	−2.057	−1.446	−0.089
	(0.081)	(0.143)	(0.107)	(0.063)
	t = −8.922***	t = −14.371***	t = −13.522***	t = −1.403
H1*–H2* (scaled)	0.280	0.077	−0.316	−0.085
	(0.080)	(0.132)	(0.102)	(0.066)
	t = 3.481***	t = 0.582	t = −3.105**	t = −1.290
H2*–H4* (scaled)	0.729	0.405	−0.126	−0.085
	(0.084)	(0.128)	(0.098)	(0.066)
	t = 8.706***	t = 3.163**	t = −1.280	t = −1.292
H4*–2kHz* (scaled)	0.253	0.026	−0.027	0.257
	(0.080)	(0.122)	(0.093)	(0.063)
	t = 3.142**	t = 0.211	t = −0.290	t = 4.068***
CPP (scaled)	−0.107	1.218	0.721	−0.023
	(0.081)	(0.120)	(0.100)	(0.068)
	t = −1.322	t = 10.137***	t = 7.181***	t = −0.340
Log Likelihood	-2239.946			
Degrees of freedom	20			
Observations	1669			
Note:				*p<0.05; **p<0.01; ***p<0.001

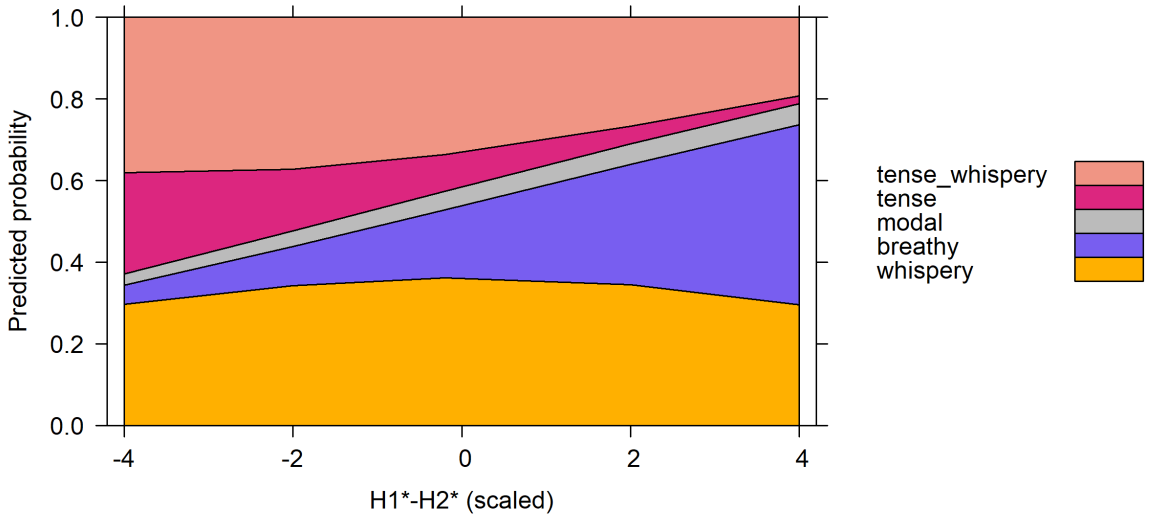


Figure 8.5: Effect of H1*–H2* on the predicted probability of a voiced stretch being rated as a particular phonation type (all other independent variables held constant at their means)

As shown in Table 8.5, the predicted probability of a voiced stretch being categorised as tense is 15% at 2 SD below the mean. This decreases as H1*–H2* increases, falling to 9% at the mean and 4% at two SD below the mean. This can be seen in Figure 8.5 by the decreasing width of the pink panel as H1*–H2* increases.

We can also be fairly certain that these effects in H1*–H2* are not due to an effect of H1*–H2* on whispery voice that cannot be seen in the model because it functions as the reference level, because the predicted probability of a voiced stretch being rated as whispery stays relatively constant, between around 34% and 36%, regardless of the value of H1*–H2*.

Scaled H1*–H2*	Raw H1*–H2* (dB)	Whispery	Breathy	Modal	Tense	Tense whispery
-2 SD	-4.81	0.34	0.10	0.04	0.15	0.37
-1 SD	-0.57	0.36	0.13	0.04	0.12	0.35
Mean (0)	3.68	0.36	0.18	0.05	0.09	0.33
+1 SD	7.92	0.36	0.23	0.05	0.06	0.30
+2 SD	12.17	0.34	0.29	0.05	0.04	0.27

Table 8.5: Predicted probabilities of a voiced stretch being rated as a particular phonation type for a range of values for H1*–H2*(all other independent variables held constant at their means)

8.3.1.2 H2*–H4* and PPA-coded phonation type

A 1-SD increase in H2*–H4* increases the log(odds) of a voiced stretch being rated as breathy rather than whispery by 0.729. This significant effect is demonstrated by the increased predicted probability of a voiced stretch being categorised as breathy as

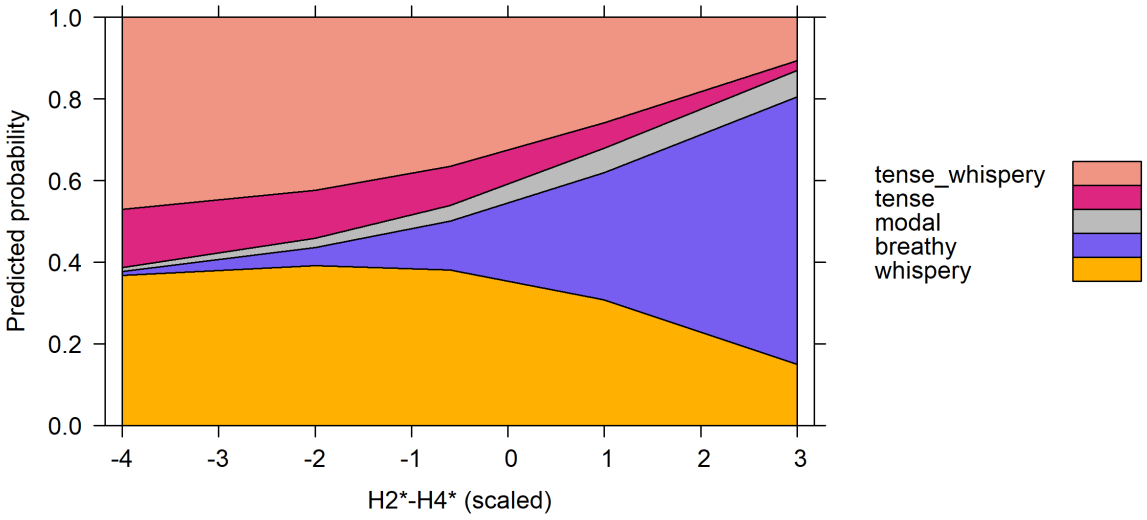


Figure 8.6: Effect of H2*–H4* on the predicted probability of a voiced stretch being rated as a particular voice quality (all other independent variables held constant at their means)

H2*–H4* increases, shown in Figure 8.6 by the increase in width in the purple panel as H2*–H4*. This effect is laid out in Table 8.6: At 2 SD below the mean, the model predicts that the probability of a breathy categorisation is 4%, while at the mean this rises to 18%, and to 48% at 2 SD above the mean.

Meanwhile, H2*–H4* also has a significant effect on the log(odds) of a voiced stretch being categorised as modal. A 1 SD increase in H2*–H4* increases the log(odds) of a voiced stretch being categorised as modal rather than whispery by 0.405. This effect can be seen in predicted probabilities of a voiced stretch being categorised as modal at different values of H2*–H4*, which increases from 2% at 2 SD below the mean, to 5% at the mean, to 7% at 2 SD above the mean.

Table 8.6: Predicted probability of a voiced stretch being rated as a particular phonation type for a range of values for H2*–H4*(all other independent variables held constant at their means)

Scaled H2*–H4*	Raw H2*–H4* (dB)	Whispery	Breathy	Modal	Tense	Tense whispery
-2 SD	-5.99	0.39	0.04	0.02	0.12	0.42
-1 SD	-1.54	0.39	0.09	0.03	0.10	0.39
Mean (0)	2.91	0.36	0.18	0.05	0.09	0.33
+1 SD	7.37	0.31	0.31	0.06	0.06	0.26
+2 SD	11.82	0.23	0.48	0.07	0.04	0.18

These effects may be attributable in part to a potential relationship between whispery voice and H2*–H4* which is not visible in the model because whispery voice functions as the reference level: As shown in Figure 8.6, the predicted probability of a voiced stretch being categorised as whispery is lower at higher values for H2*–H4*.

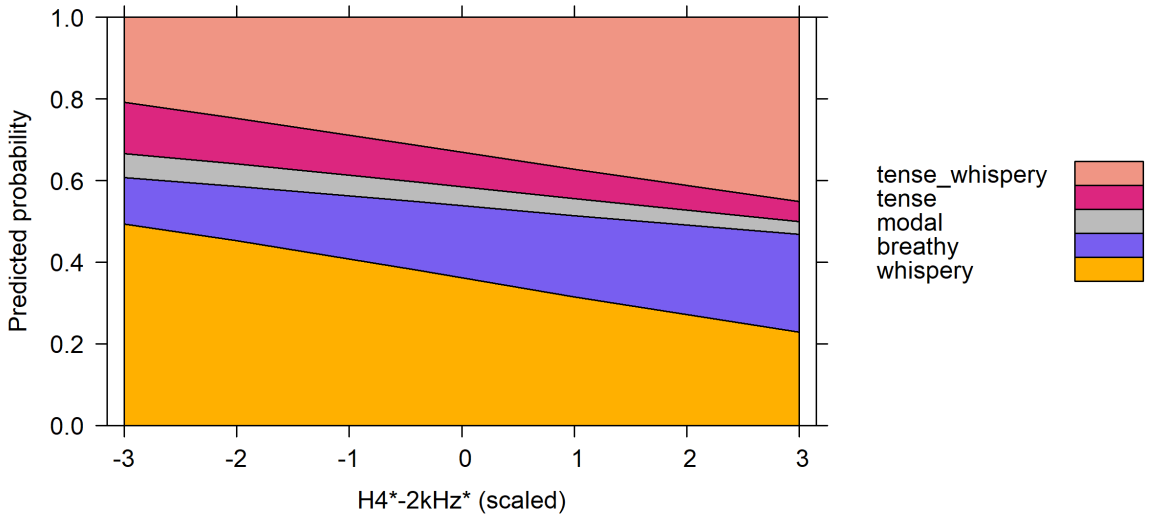


Figure 8.7: Effect of H4*–2kHz* on the predicted probability of a voiced stretch being rated as a particular phonation type (all other independent variables held constant at their means)

8.3.1.3 H4*–2kHz* and PPA-coded phonation type

H4*–2kHz* also has a significant effect on the log(odds) of a voiced stretch being categorised as breathy rather than whispery, with a 1 SD increase in H4*–2kHz* increasing the log(odds) of a voiced stretch being categorised as breathy by 0.253. This can be seen by the model predicting increased probability of a voiced stretch being categorised as breathy as H4*–2kHz* increases, from 13% at 2 SD below the mean, to 18% at the mean, to 22% at 2SD above the mean, represented by the increasing width of the purple band in Figure 8.7.

Table 8.7: Predicted probabilities of a voiced stretch being rated as a particular phonation type for a range of values for H4*–2kHz*(all other independent variables held constant at their means)

Scaled H4*–2kHz*	Raw H4*–2kHz* (dB)	Whispery	Breathy	Modal	Tense	Tense whispery
-2 SD	-9.58	0.45	0.13	0.05	0.11	0.25
-1 SD	-2.06	0.41	0.15	0.05	0.10	0.29
Mean (0)	5.46	0.36	0.18	0.05	0.09	0.33
+1 SD	12.98	0.32	0.20	0.04	0.07	0.37
+2 SD	20.50	0.27	0.22	0.04	0.06	0.41

H4*–2kHz* is the only independent variable to have a significant effect on the log(odds) of a voiced stretch being categorised as tense whispery rather than whispery: A 1 SD increase in H4*–2kHz* increases the log(odds) of a voiced stretch being rated as tense whispery rather than whispery by 0.257. This can be seen by the width of the pink-orange band increasing as H4*–2kHz* increases as in Figure 8.7. The predicted probability of a voiced stretch being categorised as tense whispery increases from 25%

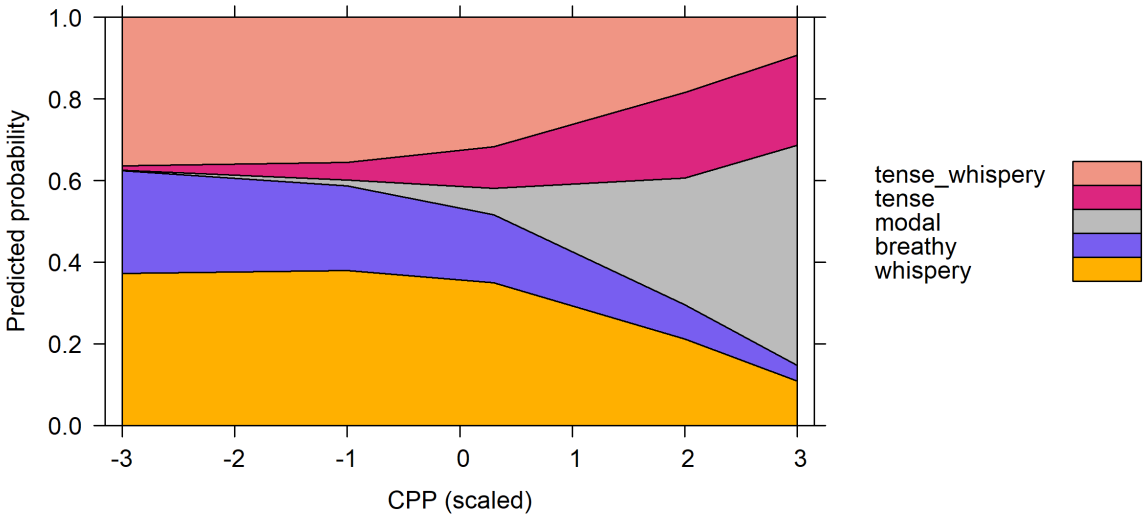


Figure 8.8: Effect of CPP on the predicted probability of a voiced stretch being rated as a particular phonation type (all other independent variables held constant at their means)

at 2 SD below the mean, to 33% at the mean, to 41% at 2 SD above the mean.

Although it is not possible to say anything about whispery voice with any certainty as it is the reference level in the model, it is possible that these effects are driven by a relationship between whispery voice and H4*–2kHz*: At 2 SD below the mean, the model predicts 45% probability of a voiced stretch being coded as whispery, while at the mean of H4*–2kHz*, it predicts 36% probability of a voiced stretch being coded as whispery, and at 2 SD above the mean, a 27% probability.

8.3.1.4 CPP and PPA-coded phonation type

Table 8.8: Predicted probabilities of a voiced stretch being rated as a particular phonation type for a range of values for CPP (all other independent variables held constant at their means)

Scaled CPP	Raw CPP (dB)	Whispery	Breathy	Modal	Tense	Tense Whispery
-2 SD	15.49	0.38	0.23	0.00	0.02	0.36
-1 SD	18.04	0.38	0.21	0.01	0.04	0.36
Mean (0)	20.59	0.36	0.18	0.05	0.09	0.33
+1 SD	23.14	0.31	0.13	0.13	0.15	0.28
+2 SD	25.69	0.21	0.08	0.31	0.21	0.18

CPP was the only independent variable not to have a significant effect on the log(odds) of a voiced stretch being coded as breathy rather than whispery.

However, CPP significantly affected the log(odds) of a voiced stretch being coded as modal: A 1 SD increase in CPP, which represents a *decrease* in noise, increases the log(odds) of a stretch being coded as modal by 1.218. This is demonstrated by the

model predicting a 0% probability of a voiced stretch being modal at -2 SD below the mean for CPP, a 9% chance at the mean of CPP, at a 31% chance at 2 SD above the mean. This is represented by the grey band increasing as CPP increases in Figure 8.8.

There was also a significant effect of CPP on the log(odds) of a voiced stretch being coded as tense voice, which increases by 0.721 as CPP increases by 1 SD. The predicted probability of a voiced stretch being coded as tense voice, shown by the pink band in Figure 8.8, increases from 2% at 2SD below the mean, to 9% at the mean of CPP, to 21% at 2SD above the mean.

8.3.2 Analysis of the scalar degrees of whispery voice

As whispery voice was the most commonly occurring simple phonation type (Figure 6.12), I looked more closely at how the scalar degree of whispery voice varied as a function of CPP.

Table 8.9 shows the results of the multinomial model predicting the degree of whispery voice as a function of CPP.

Table 8.9: Summary of multinomial logit model predicting the degree of whispery voice for whispery voiced stretches as a function of CPP. In each cell the estimate is presented first in log-odds, followed by the standard error (SE) in brackets, then the t-value and any asterisks indicting statistical significance.

	Scalar degree:		
	2	3	4
Constant	0.899	1.168	−0.463
	(0.157)	(0.152)	(0.225)
	t = 5.731***	t = 7.688***	t = −2.061*
CPP (scaled)	−0.603	−1.011	−2.121
	(0.153)	(0.156)	(0.221)
	t = −3.946***	t = −6.497***	t = −9.600***
Log Likelihood	-644.1227		
Degrees of freedom	6		
Observations	548		
Note:			*p<0.05; **p<0.01; ***p<0.001

As CPP increases by 1 SD, representing a decrease in noise, the log(odds) of a whispery voiced stretch being rated as scalar degree 2 rather than scalar degree 1 decrease by 0.603, and the log(odds) of a whispery voiced stretch being rated as scalar degree 3 rather than scalar degree 2 decrease by 1.011, and the log(odds) of a whispery voiced stretch being rated as scalar degree 4 rather than scalar degree 1 decrease by 2.121.

This effect is shown in Figure 8.9, which shows how the model predicts a lower probability of higher scalar degrees of whispery voice as CPP increases (=as noise decreases), represented in the plot by the shift from more area being occupied by darker orange at lower values for CPP to more area being occupied by lighter yellow for higher values for CPP. Table 8.10 lays out how these predicted probabilities vary as a function of CPP: At low values for CPP, 2 SD below the mean, the model predicts a 1% probability of a stretch being rated scalar degree 1, an 11% probability for scalar degree 2, a 31% probability for scalar degree 3, and a 57% probability for scalar degree 4. At high values for CPP, 2 SD above the mean, the model predicts a 46% probability

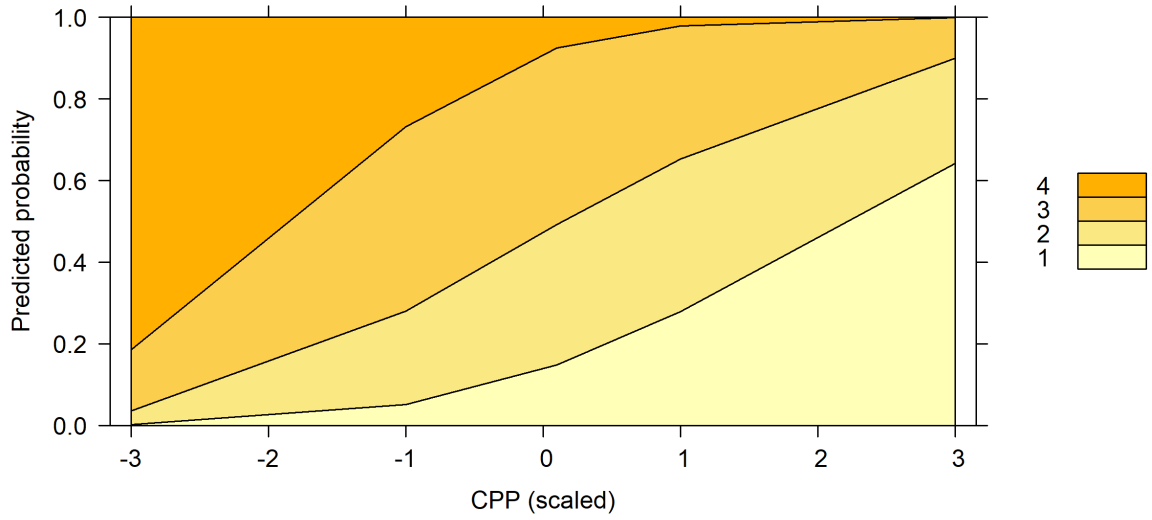


Figure 8.9: Effect of CPP on the predicted probability of a voiced stretch being rated as each scalar degree of whispery voice

of scalar degree 1, a 34% probability of scalar degree 2, a 20% probablity of scalar degree 3, and a 0% probability of scalar degree 4.

Scaled CPP	Raw CPP	1	2	3	4
-2 sd	15.49	0.01	0.11	0.31	0.57
-1 sd	18.04	0.05	0.23	0.45	0.27
Mean	20.59	0.14	0.34	0.44	0.09
+1 sd	23.14	0.28	0.37	0.33	0.02
+2 sd	25.69	0.46	0.34	0.20	0.00

Table 8.10: Predicted probabilities of a whispery voiced stretch being rated as each scalar degree for a range of values for CPP (all other independent variables held constant at their means)

8.3.3 Analysis of scalar degrees of breathy voice

Breathy voice contained the second highest number of voiced stretches coded as exclusively as a single phonation type. Because of this, I looked more closely within the category of breathy voice, and analysed how scalar degree of breathy voice varied as a function of H1*–H2*, H2*–H4*, and CPP. The results of the multinomial model analysing this are presented in Table 8.11.

	Scalar degree:		
	2	3	4
Constant	−0.033	0.308	−1.501
	(0.176)	(0.164)	(0.327)
	t = −0.186	t = 1.879	t = −4.591***
H1*–H2* (scaled)	0.024	0.366	0.458
	(0.174)	(0.168)	(0.218)
	t = 0.140	t = 2.183*	t = 2.103*
H2*–H4* (scaled)	0.170	0.427	0.574
	(0.168)	(0.166)	(0.220)
	t = 1.015	t = 2.565*	t = 2.610**
CPP (scaled)	−0.443	−1.077	−2.467
	(0.170)	(0.188)	(0.350)
	t = −2.602**	t = −5.734***	t = −7.046***
Log likelihood	-382.4009		
Degrees of freedom	12		
Observations	335		
Note:			*p<0.05; **p<0.01; ***p<0.001

Table 8.11: Summary of multinomial logit model predicting the scalar degree of breathy voice for breathy voiced stretches as a function of H1*–H2*, H2*–H4*, and CPP. In each cell the estimate is presented first in log-odds, followed by the standard error (SE) in brackets, then the t-value and any asterisks indicting statistical significance.

In the model analysing phonation type as a function of acoustic measures, summarised in Table 8.4, H1*–H2* had a significant effect on the log(odds) of a stretch being coded as breathy rather than whispery. Here, H1*–H2* also has some significant effects on the scalar degree of breathy voice: While there is no effect on the log(odds) of a stretch being coded as scalar degree 2 rather than 1, a 1 SD increase in H1*–H2* increases the log(odds) of a stretch being coded as scalar degree 3 rather than scalar degree 1 by 0.366. Furthermore, a 1 SD increase in H1*–H2* increases the log(odds) of a stretch being coded as scalar degree 4 rather than scalar degree 1 by 0.458.

This is demonstrated by the change in predicted probabilities of each scalar degree as H1*–H2* increases, shown in Table 8.12 and Figure 8.10

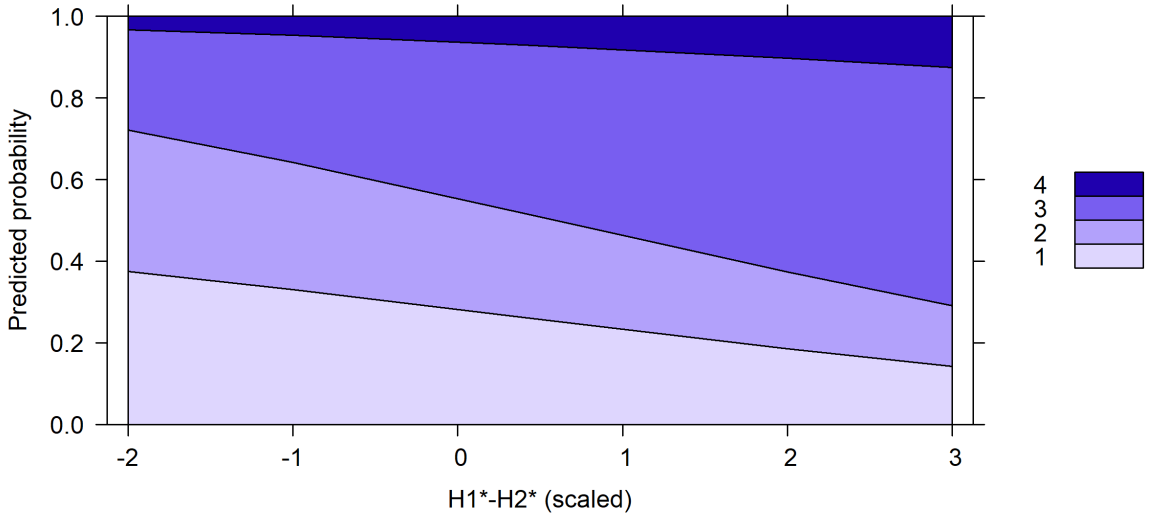


Figure 8.10: Effect of H1*–H2* on the predicted probability of a voiced stretch being rated as a higher scalar degree of breathy voice

Table 8.12: Predicted probabilities of a breathy voiced stretch being rated as each scalar degree for a range of values for H1*–H2*(all other independent variables held constant at their means)

Scaled H1*–H2*	Raw H1*–H2*	1	2	3	4
-2 sd	-2.99	0.38	0.35	0.25	0.03
-1 sd	0.92	0.33	0.31	0.31	0.05
Mean	4.82	0.28	0.27	0.38	0.06
+1 sd	8.73	0.23	0.23	0.46	0.08
+2 sd	12.64	0.19	0.19	0.52	0.10

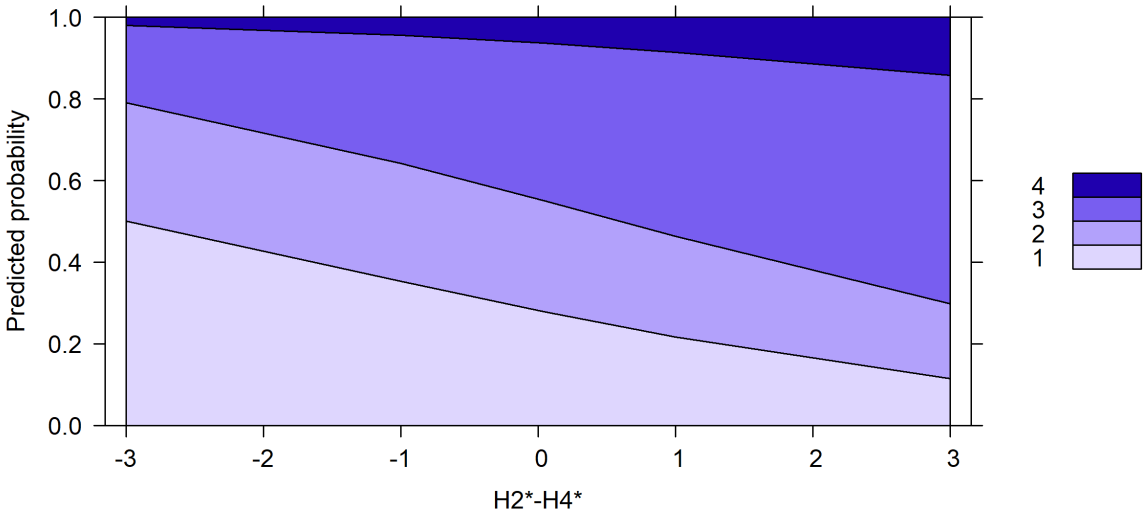


Figure 8.11: Effect of H2*–H4* on the predicted probability of a voiced stretch being rated as a higher scalar degree of breathy voice

H2*–H4* also had a significant effect on the log(odds) of a stretch being coded as breathy rather than whispery in the model analysing phonation type as a function of various acoustic measures. Here, H2*–H4* also has a significant effect on the scalar degree of breathy voice. Although it does not have an effect on the log(odds) of a stretch being coded as scalar degree 2 over scalar degree 1, a 1 SD increase in H2*–H4* increases the log(odds) of a stretch being coded as scalar degree 3 rather than 1 by 0.427, and of a stretch being coded as scalar degree 4 rather than 1 by 0.574.

This is demonstrated by the change in predicted probabilities of each scalar degree as H2*–H4* increases, shown in Table 8.13 and Figure 8.11

Table 8.13: Predicted probabilities of a breathy voiced stretch being rated as each scalar degree for a range of values for H2*–H4*(all other independent variables held constant at their means)

Scaled H2*–H4*	Raw H2*–H4*	1	2	3	4
-2 sd	-2.42	0.43	0.29	0.25	0.03
-1 sd	1.48	0.35	0.29	0.31	0.04
Mean	5.38	0.28	0.27	0.38	0.06
+1 sd	9.29	0.22	0.25	0.45	0.09
+2 sd	13.19	0.16	0.22	0.51	0.11

In the model that analysed voice quality as a function of different acoustic measures, an increase in CPP did not increase the log(odds) of a stretch being coded as breathy rather than whispery. However, within breathy voice, an increase in CPP does decrease the log(odds) of a stretch being coded as a higher scalar degree of breathy voice. A 1 SD increase in CPP decreases the log(odds) of a stretch being coded as scalar degree 2 rather than scalar degree 1 by 0.443, and of a stretch being coded as scalar degree 3 rather than 1 by 1.077, and of scalar degree 4 rather than 1 by 2.467. This effect is

shown by the changes in predicted probability given in Table 8.14 and shown in Figure 8.12.

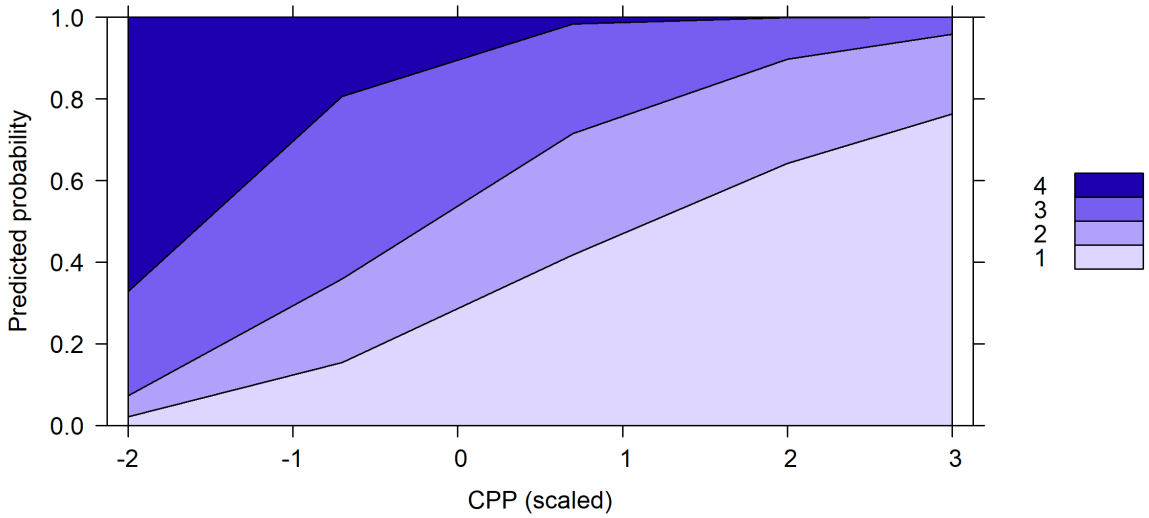


Figure 8.12: Effect of CPP on the predicted probability of a voiced stretch being rated as a a higher scalar degree of breathy voice

Table 8.14: Predicted probabilities of a breathy voiced stretch being rated as each scalar degree for a range of values for CPP(all other independent variables held constant at their means)

Scaled CPP	Raw CPP	1	2	3	4
-2 sd	14.66	0.02	0.05	0.25	0.67
-1 sd	17.26	0.11	0.17	0.44	0.29
Mean	19.86	0.28	0.27	0.38	0.06
+1 sd	22.46	0.48	0.30	0.22	0.01
+2 sd	25.06	0.64	0.26	0.10	0.00

8.4 Discussion: Connecting PPA and acoustic analysis of non-creaky voice

I conducted a Phonation Profile Analysis of 24 speakers across Glasgow, Lothian and Shetland, including two male and two female speakers from each age group from each area. I then compared my auditory coding of creak using PPA to the creak coded using an automated f0-based method to investigate places where coding diverged. I undertook an acoustic investigation of non-creaky voice by measuring spectral tilt and noise measurements in each voiced stretch. I then conducted a between-category logistic regression model that investigated the relationship between auditorily coded voice quality and H1*–H2*, H2*–H4*, H4*–2kHz* and CPP, and two within-category models that modelled the scalar degree of breathy voice as a function of H1*–H2*, H2*–H4*, and CPP, and the scalar degree of whispery voice as a function of CPP. Here, I discuss the ways that voice quality manifested acoustically and what this suggests about the glottal configurations underlying these different qualities. I then turn to the ways that voice quality varies according to social factors in PPA of this same data, and formulate predictions about how acoustic measures of voice quality may vary according to social factors in the larger corpus study.

8.4.1 Whispery voice

I expected that the high noise present whispery voice would manifest acoustically as low CPP. I also expected that the long open phase of whispery voice would give it a moderately steep slope for H1*–H2* and H2*–H4*, and that the posterior glottal gap would create a steep slope for H4*–2kHz*. I also expected that as the degree of whispery voice increased, CPP would lower, and slope would increase.

H1*–H2* and H2*–H4* both relate to glottal openness and the open phase, where higher values suggest a more open configuration. In the data, spectral tilt for whispery voice appears to have a tendency to have a moderate spectral slope. This can be seen firstly in the raw values for spectral tilt of whispery voice. The mean and median value for H1*–H2* in whispery voice is 3.8 dB (max=18.8, min=–11.9, Q1=1.5, Q3=6). Compared to previous studies, whispery tokens here appear to span a wide range of values from what would usually be considered evidence of creaky voice (–11.9 dB), to what would often be considered very breathy (18.8 dB), while the middle 50% of the data appear to fall in a range that could be taken to be evidence for modal or moderately breathy voice. For example, Henton & Bladon's (1985) study of breathy voice in English accents attributes mean differences between the first and second harmonics of between 3.3 to 8.4 dB in female speakers to breathy voice and differences of 0–1 dB in male speakers to a lack of breathy voice. In Bickley's (1982) study of phonemic phonation

contrasts, she finds a mean difference between H1 and H2 of 7 dB for breathy vowels in !Xóõ (ranging from –4 to 15 by speaker), and between 0 and –9 for 'clear' vowels in data collected by Ladefoged (1971). Furthermore, she finds that !Xóõ vowels judged to be 'very breathy' by Gujurati listeners have an H1–H2 difference of 5.3–12.5 dB, while vowels judged to be non-breathy had an H1–H2 difference of 0. Overall, whispery voice in this study tends to occupy a space between modal and breathy voice in terms of H1–H2, while overlapping with both modal and breathy values at its extremes.

H2*–H4* is less widely used in phonetic research. Garellek, Ritchart & Kuang (2016) model relationships between different spectral tilt measures and find a range of –5.1 to 29.2 dB for H2–H4 among speakers with and without voice disorders, with a mean of 11.5 dB for female speakers and 8.9 dB for male speakers. Here, H2*–H4* ranges from –5.3 dB to 16.4 dB, with a mean and median of 5.4 dB (Q1=2.8, Q3=8.1). In comparison, the range for H2*–H4* for whispery voice here appears to extend less high, and seems to have a lower mean. This could suggest that whispery voice here is produced with a less open glottis than much voice more generally is.

Comparing spectral tilt measures to the values found in previous research is limited in usefulness as different research may take differing approaches to conceptualising and measuring voice quality. Instead, it is perhaps more useful to note that in terms of H1*–H2*, whispery voice here is not significantly different from modal voice and lies between breathy voice and tense voice, but is significantly lower than breathy and modal voice in terms of H2*–H4*, patterning alongside tense voice. This suggests that whispery voice may be produced with high vocal fold contact and a short open phase, but in a way that boosts H4 rather than H2. This is in line with values for H2*–H4* tending towards the lower end of the range presented by Garellek, Ritchart & Kuang (2016).

H4*–2kHz* has a relationship with the abruptness of vocal fold closure, and is particularly relevant for whispery voice because posterior the glottal gap that is present in whispery voice should theoretically increase the slope for H4*–2kHz*. Garellek, Ritchart & Kuang (2016) find values for H4*–2kHz* ranging from 2–43.2 dB, with a mean of 18.1 dB for female speakers and a mean of 24.6 for male speakers. Here, whispery voice has a mean and median of 6.6 dB for H4*–2kHz*, and ranges from –14.9 to 24.6 (Q1=1.9, Q3=11.6). Whispery voice lies between values for breathy and tense whispery voice, perhaps suggesting that in this study, it is not characterised by a posterior glottal gap. However, it could also mean that H4*–2kHz* may be picking up noise rather than harmonic energy at 2kHz. Spectral measurements work by first finding f0, then using this to locate harmonics, and and measuring amplitude in the region where the formula *expects* the harmonic to be: If it looks for a harmonic in the region of 2kHz, but a voiced stretch is very whispery, this could lead to it taking a local peak of energy as the harmonic when it was actually noise. For example, Gobl &

Chasaide (1992) found additional noise in whispery voice in the upper regions of the spectrum when looking at the acoustic profile of different voice qualities in the LTAS.

I expected whispery voice to be characterised by high noise, manifesting here as low CPP. I found that CPP for whispery voice had a mean of 19.9 dB and a median of 19.6 dB, and varied between 14.7 and 28.6 (Q1=18, Q3=21.2). Modal voice and tense voice both showed significantly higher CPP than whispery voice, which suggests in turn that whispery voice is characterised by lower CPP (i.e. increased aperiodic noise) than modal and tense voice. As the degree of whispery voice increases in a whispery voiced stretch, CPP is also the only measure to change significantly. This suggests that whispery voice may be characterised more by noise than by spectral tilt. This is in line with expectations, and follows previous work by Tian & Kuang (2021) who looked at different phonemic phonation types in different languages and found that in contrast to breathy voice in other languages, whispery voice in Shanghaiese is characterised more by noise than by spectral tilt.

The aperiodic noise that characterises whispery voice relative to other qualities and increasing auditory degrees of whispery voice, along with H4 potentially being boosted, suggests that the laryngeal constrictor plays a key role in the production of aperiodic noise in whispery voice here, rather than incomplete closure of the vocal folds.

8.4.2 Breathy voice

I expected that breathy voice would be characterised by steep spectral slope, steeper than that of whispery voice. I posited that breathy voice might have a less steep slope than whispery voice in the H4*–2kHz* region due to incomplete closure along the length of the vocal folds rather than the 'Y-shaped' configuration that is typical of whispery voice.

The model comparing different voice quality categories showed that breathy voice exhibited a steeper spectral slope than whispery voice, which was particularly apparent for H2*–H4*. This suggests that in comparison to whispery voice, breathy voice is characterised by a less complete closure, longer open phase, and more gradual cut-off between the closed and open phases.

I expected that whispery voice would be characterised more strongly by noise, while breathy voice would be characterised more strongly by spectral tilt. The findings here confirm this: Both noise and spectral tilt are important components of whispery voice and breathy voice, but breathy voice involves increased spectral tilt compared to whispery voice, and increasingly breathy qualities involve increasing spectral tilt, while increasing spectral tilt plays no significant role in the auditory percept of increased whisperiness.

8.4.3 Modal voice

I expected that modal voice would have a flatter spectral slope than whispery voice. Contrary to this, I found no significant difference between whispery and modal voice for H1*–H2* or for H4*–2kHz*, and a greater slope for H2*–H4*.

I also expected that modal voice would be less noisy than whispery voice. Consistent with this prediction, I found that modal voice had a significantly higher CPP than whispery voice, indicating that it contained less aperiodic noise.

The similarities in spectral slope in terms of H1*–H2* and H4*–2kHz* could suggest that whispery and modal voice are produced with similar glottal configurations. However, the fact that modal voice is much less noisy suggests that whispery voice is produced with epilaryngeal constriction, creating a turbulent airflow, while modal voice is not. If the production underlying whispery voice boosts H4, increased H2*–H4* in modal voice may reflect efficient vocal fold vibration in comparison to whispery.

8.4.4 Tense voice

I expected that tense voice would have a less steep spectral slope than whispery voice. I found this in terms of H1*–H2*, suggesting that tense voice has a shorter open phase or is more constricted than whispery voice. Contrary to expectations, I found no difference in H2*–H4* or H4*–2kHz* between whispery and tense voice. This suggests that tense voice may have a shorter open phase than whispery voice, but that tense voice does not involve a sharper glottal closure than whispery voice.

I also expected that tense voice would be less noisy than whispery voice. I found that CPP was higher for tense voice, indicating that tense voice is produced with less aperiodic noise than whispery voice.

8.4.5 Tense whispery voice

I expected that tense whispery voice might pattern similarly to whispery voice, but might lie in between tense and whispery voice or exhibit some features that are typical of tense voice.

I found no significant difference between tense whispery voice and whispery voice in terms of H1*–H2* or in terms of H2*–H4*, but greater H4*–2kHz*. This might suggest that tense whispery voice is produced with a larger glottal gap than whispery voice, leading to a more gradual glottal closure. However, there are several other possibilities for greater H4*–2kHz*. One is that tense whispery voice and whispery voice may

differ in terms of the frequencies that contain aperiodic noise, and that whispery voice may have more noise in the 2kHz region which is picked up as spectral energy during measurement, leading to a flatter slope between H4* and 2kHz for whispery voice, and by comparison a steeper slope for tense whispery voice.

8.4.6 Conclusions

While some auditory voice quality categories, such as tense voice, manifested acoustically in a way that was very close to what was expected, others did not manifest acoustically in the way that might be expected from what we assume about the vocal fold configuration that underlies their production. For example, the posterior glottal gap that is present in the production of whispery voice would be expected to produce a steep spectral slope in the H4*-2kHz* region, but I find that whispery voice appears to have tendency towards a fairly flat spectral shape in the upper regions of the spectrum, while tense whispery voice does exhibit a steep slope in this region. These unexpected findings suggest that speakers may not be using the expected configuration to produce each voice quality, and leading to an absence of a one-to-one mapping between the PPA ratings and acoustic measurement.

This is supported by the fact that spectral tilt measures also tended to span a wide range within each voice quality category, especially for whispery and tense whispery voice. This suggests that speakers may be using different vocal fold configurations to produce similar auditory effects. Furthermore, interspeaker differences may help account for this as well, as different speakers will be using different vocal anatomies to produce voicing, and may use different strategies to account for this. The sample considered here includes male and female speakers, who may be using different strategies to achieve the auditory effects of different voice qualities. Furthermore, I did not account for the effect of nasal poles in the data, which would have affected the amplitude of harmonics differently at different fundamental frequencies (Simpson 2012), and led to different tendencies for spectral tilt measures in nasalised vowels and nasals. Furthermore, I included both speakers aged between 18–25 and speakers who were over 65; the vocal anatomy of older speakers may have been affected by ageing, habits such as smoking through the lifespan, and other factors. This in turn could have caused different acoustic profiles for tokens with similar auditory voice quality in older and younger speakers. The multinomial models used here did not contain a random effect for speaker, so would not be able to control for this phenomenon.

Together, these findings together suggest that there may not be one-to-one mapping between the auditory-perceptual voice quality categories and differences in acoustic measures, and that this may arise from different tokens that are categorised within the same category being produced with different underlying configurations.

Despite the fact that not all voice qualities manifested acoustically in the way that I expected, each category and increasing degrees within category can be differentiated from each other on the basis of one or more acoustic measures. Of particular note is the fact that although breathy and whispery voice share some acoustic similarities, they also show differences. Firstly, in the comparison of different voice qualities, they shared high noise levels, but could be differentiated on the basis of spectral tilt, which was steeper for breathy voice at all levels. Furthermore, the within-category analyses revealed that higher degrees of breathy voice were characterised by both higher noise and increased spectral tilt, while higher degrees of whispery voice were only characterised by higher noise. In the analysis comparing different voice quality categories, modal voice, tense voice and tense whispery voice also differed from whispery voice on at least one acoustic measure.

Looking forward to the larger corpus analysis, these results suggest that it should be possible to use these measures of voice quality in voiced stretches that come from spontaneous speech data collected in non-lab settings. Furthermore, the present analysis means that it should be possible to interpret the results of the subsequent analysis with reference to traditional labels for voice quality, allowing comparison between this research and previous VPA studies of voice quality in Scotland. In Chapter 9, I present an analysis of how voice quality varies according to social and linguistic factors in 95 speakers from SCOSYA, using 180 seconds of speech from each speaker. In Section 9.1, I begin by bringing together the results of the PPA and the analyses that compared PPA and acoustic analysis and f0-based automatic detection of creak, in order to set out some predictions for how the acoustic measures might pattern in the analysis in the wider corpus.

Chapter 9

Scaling up: Acoustic analysis of voice quality in Scottish accents in a larger sample

9.1 Introduction

In this chapter, I consider how voice quality varies according to social and linguistic factors in Scottish accents in an acoustic analysis of 95 speakers from SCOSYA. This analysis involves more data from each speaker than the smaller sample used in the analyses thus far, with 180 seconds of speech taken from each speaker rather than the 90 seconds used in the smaller sample. This 95 speaker sample includes participants from Glasgow, Lothian (Edinburgh and surrounding areas) and both Orkney and Shetland (Insular Scots), stratified by gender (49 female, 46 male), age (49 aged 18-25, 46 aged 65+) and area (19 Insular, 28 Lothian, 48 Glasgow).

To investigate the social and linguistic variation in voice quality in this data, I build on the methods used thus far and take a two-stage approach. First, I use f0-based coding of creak to separate out creaky voice from non-creaky voice and estimate how use of creak varies according to social and linguistic factors. Then, I consider non-creaky voice in more detail, and investigate how H1*–H2*, H2*–H4*, H4*-2kHz* and CPP vary according to social and linguistic factors. Having separated out creaky and non-creaky voice then allows a more straightforward interpretation of variation in voice quality in the analysis.

9.1.1 Predictions

9.1.1.1 Social factors

In the PPA analysis, voice quality in Glasgow appeared to mostly be characterized by use of whispery and tense whispery voice. If this pattern is seen in the larger corpus, I would expect to find voice quality in Glasgow having low CPP, low H2*-H4*, and high H4*-2kHz*. I also expect to find low levels of f0-based creaky voice in Glasgow speech.

Voice quality in Lothian appeared to be characterized by modal and near-modal voice quality such as tense voice, as well as use of breathy voice and occasional use of harsh voice in some speakers. In Lothian speakers, I therefore expect to find increased H2*-H4*, reflecting breathy and modal voice quality, but to find high H1*-H2* and low CPP where this formed part of breathy voice, and high CPP where this formed part of modal voice. I would also expect to find some cases of high CPP and low H1*-H2*, reflecting tense voice, as well as of extremely low CPP, reflecting harsh voice. I also expect to find low levels of f0-based creaky voice in Lothian speech, some of which may in fact be cases of harsh voice.

In Insular voice quality, I expect to find more creaky voice, which I expect will mostly be identified in the f0-based system. In terms of acoustic measures, I expect

Insular varieties to sit in-between Glasgow and Lothian. I also hypothesise that it is likely that a more distinct acoustic profile will be revealed in the acoustic analysis of Insular voice quality than was revealed in the PPA analysis, because of the way that acoustic measurement allows a more fine-grained analysis of variation that is difficult to capture using labels and ordinal scales. It could be, for example, that the specific configuration used to produce a 'breathier' voice quality in Insular varieties is acoustically distinct from the whispery voice quality used in Glasgow and the breathy voice quality used in Lothian, but that this is not captured in the PPA labelling system.

I expect to find some broad differences between male and female speakers generally. I expect that male speakers will use voice qualities that are nearer to modal voice, and therefore use higher CPP and higher H2*–H4*. However, I expect that where male speakers use breathy voice they will use breathier breathy voice, so that where male speakers demonstrate evidence of breathy voice in terms of higher H1*–H2* and lower CPP, they will use higher H1*–H2* and lower CPP than female speakers. I also expect that these differences may not be consistent across areas and age groups, in a way that I have not yet been able to capture in the smaller corpus, because of the small number of speakers in each cell for each age/gender/area category.

Among female speakers, I expect to find lower CPP than in male speakers overall, reflecting a more whispery voice quality, as well as low H1*-H2* with high CPP, reflecting tense voice. Where female speakers use higher H1*–H2*, indicating a breathy quality, I expect this will be accompanied by a higher CPP than in male speakers, indicating a more lax-style, 'less breathy' use of breathy voice.

In the PPA analysis, I did not find many differences between older and younger speakers in the types of voice quality used, but I did find that older speakers used less creaky voice, less modal voice, and more harsh voice. Because of this, I firstly expect to find that younger speakers use more creak than older speakers in the larger corpus. Furthermore, based on decreased rates of modal voice, increased harsh voice, and previous research on the effects of ageing on the voice (Rojas, Kefalianos & Vogel 2020), I expect that CPP will be lower among older speakers, reflecting increased noise. I also expect that there may be differences between older and younger speakers that will differ by gender and area which were not visible in the smaller corpus because of the low number of speakers in each cell; Research that considered gender variation in ageing effects on the voice tentatively finds that while male voices tend to show increased noise with age, this finding is less well-established for female speakers who may decreased spectral tilt with age (Gittelson, Leemann & Tomaschek 2021, Lee et al. 2016, Rojas, Kefalianos & Vogel 2020). I also think it is possible that differences will emerge between older and younger speakers which were not captured by PPA, as similar phonation types produced by speakers of different age groups may manifest differently acoustically, and the continuous nature of acoustic methods may be better

suited to capture this.

I also expect there may be differences between older and younger speaker in Insular voice quality. This is because sociolinguistic work has found that age is an important factor in speakers' use of phonetic, morphosyntactic and lexical features, with a sharp decline in local forms used by younger speakers (Smith & Durham 2011).

9.1.1.2 Linguistic factors

In the smaller corpus, the two linguistic factors considered, phrase-final position and glottal context, both favoured the presence of creak as expected. As detailed in Section 9.2, I again used the voiced stretch as the unit of analysis, but did not hand-correct stretches to exclude the influence of surrounding segments, so more linguistic factors can be expected to have an effect here. I therefore used the output of the forced aligner to estimate the influence of linguistic context on voice quality. Here, I detail how I expect voice quality to vary according to linguistic factors, drawing on the research summarised in Section 5.2.4.

9.1.1.2.1 Duration In the smaller corpus, stretches below 100ms were excluded as an arbitrary threshold below which making auditory-perceptual judgements about voice quality became more difficult. In the larger corpus, I am not making auditory-perceptual judgements, but still want to control for potential effects of stretch duration. I measured the duration of the stretch. I expected that shorter stretches were more likely to include more noise from surrounding segments, so expected that shorter stretches would be noisier and show lower values for CPP.

9.1.1.2.2 Speech rate I measured speech rate in syllables per second. Little research considers the effect of speech rate on voice quality (Though see Pratt (2023) for a study which found that bodily construction of 'chill' affect was co-constructed by slower speech rate and use of creak), so I did not have any specific expectations on how I expected speech rate to affect incidence of creak and acoustic measures in non-creaky voice quality. However, Gerratt, Kreiman & Garellek (2016: 999) note speech rate as one aspect that might be important to consider if aiming to quantify variation in voice quality in continuous speech.

9.1.1.2.3 Glottalisation The potential effects of glottal stops on voice quality has already been explored in the smaller corpus, where as predicted, they increased the amount of tense voice and creak (See Section 6.3.3). However, while in the smaller corpus I hand-coded the presence of glottal stops, in the larger corpus I will be esti-

mating the potential effect of glottals from the output of the forced aligner. As such, I define 'glottalable context' (termed such because these contexts are possible to glottal, but not certainly glottaled) in the contexts that Cruttenden & Gimson (2014: 184) outlines where a /t/ could potentially be replaced by a glottal stop, which also apply to Scottish English. Criteria for glottalable contexts are as follows:

1. The /t/ occurs in syllable-final position
2. The /t/ is followed by one of the following:
   - a pause (e.g. *bet #* [bɛʔ])
   - a vowel (e.g. betting [bɛʔɪŋ], bet on [bɛʔ ɔn]),
   - /n/, /m/, or /l/ (e.g. *button* [bʊʔn̩], *little*[lɪʔl̩]
3. The /t/ is preceding by one of the following:
   - a vowel (e.g. *bet #* [bɛʔ])
   - /l/ (e.g. *belt* [bɛlʔ])
   - /n/ (*bent bɛnʔ*
   - /r/
   - *part* (e.g. [paɹʔ])

I also consider potential pre-glottalisation contexts are likely to favour creak and tense voice. Following Cruttenden & Gimson (2014: 184), I define these as cases where a stretch is followed by /p/, /t/, /k/ or /tʃ/, which itself meets the following criteria:

1. It occurs at the end of a syllable
2. In the case of /p,t,k,/, it must be followed by a pause (e.g. *think #*, [θɪŋk] #, or a consonant (e.g. *think that*, [θɪŋk ðat])
3. In the case of /tʃ/, it can be followed by any sound (e.g. *catching*, [katʃɪŋ])

9.1.1.2.4 Aspiration As mentioned briefly in Section 5.2.4, aspiration noise from preceding obstruents may affect voice quality. As Ladefoged (2006: 56-57) notes, English voiceless plosives /p,t,k/ are produced with aspiration, a period of voicelessness after the release of the plosive before the onset of voicing of the following vowel, unless they are preceded by /s/. The presence of a preceding voiceless plosive can also lead to approximants to become devoiced, as in *play* realised as [pl̥e]. I expect that some of these periods of voicelessness and devoicing may be included in the voiced stretches identified from the output of the forced aligner, and that in turn aspiration

from preceding voiceless aspirated plosives may cause increased noise (=lower CPP) and spectral tilt (=higher H1*–H2*, H2*–H4*, H4*–2kHz). I expect similar findings following a voiceless glottal fricative /h/, which as Ladefoged (2006: 265) notes, can be understood as a voiceless transition into a syllable.

9.1.1.2.5 Pre-aspiration As outlined in Section 5.2.4, I expect that pre-aspiration may also have an effect, as it manifests as a period of voicelessness or breathier voicing so can be expected to increase noise and spectral tilt. Pre-aspiration can occur in vowels that precede non-word-initial voiceless fricatives (Hejná 2015, Hejná & Scanlon 2015, Gordeeva & Scobbie 2010); for example, at the end of the vowel in the word *grass*.

9.1.1.2.6 Phrase-final position As found in the smaller corpus in Section 6.3.3, I expect that occurring in phrase-final position will increase the chance of creak occurring. I also expect that acoustic cues to increased glottalisation will occur; that is, decreased spectral tilt.

9.1.1.2.7 Phrase-initial vowels As outlined in Section 5.2.4, glottalisation is likely in the case of vowels that occur at the beginning of a phrase. I therefore expect increased creak and acoustic cues to glottalisation will occur in these contexts.

9.1.1.2.8 Phones contained in the stretch I expect that other phones contained in the stretch may also affect the acoustic measures of voice quality, though make no particular predictions as to how, given that most previous research on voice quality operates exclusively on vowels.

9.2 Methods

9.2.1 Corpus

For the wider acoustic analysis in SCOSYA, more speakers were selected from SCOSYA from the same three dialect areas as in the smaller corpus. However, as there were fewer speakers from Shetland, Orkney speakers were also considered for inclusion under the umbrella of 'Insular Scots' speakers. All speakers from Glasgow, Lothian and Orkney & Shetland were considered for inclusion in the expanded corpus. SCOSYA metadata identifies a total of 205 speakers from these areas, of which 85 are from Glasgow, 94 from Lothian, and 26 from Orkney & Shetland. Initially my intention had also been to include speakers from two other Scots varieties, North-East and the Western Isles, to compare five Scots dialect areas, however this was not possible due to time constraints.

9.2.1.1 Recording screening

Speakers were then selected for inclusion on the basis of an auditory assessment of the level of background noise and recording quality. Recordings where the speaker was distant from the microphone or where there was a high level of persistent background noise were excluded to improve the results of the forced alignment and subsequent F0 tracking. This resulted in the selection of 66 speakers from Glasgow, 19 from Orkney & Shetland, and 40 from Lothian, reducing the sample to 125 speakers. 180 seconds of speech were extracted from each speaker, using Script 3, which selected the longest turn of the speaker and the preceding or following speech. This was a similar process to how speech was selected in the smaller corpus, but with more speech extracted per speaker.

9.2.1.2 Forced alignment and hand-check for background noise

Time-aligned transcriptions were provided as part of SCOSYA data. These were converted to TextGrids using a script (Fromont 2022), then force aligned using the Montreal Forced Aligner v.1.1 (McAuliffe et al. 2017b,a). Each continuous stretch of speech was hand-checked for local background noise, constructed dialogue, and major forced alignment errors, which were excluded using Script 4. Minor errors that were straightforward to correct, such as the location of boundaries at the start and end of utterance, were corrected at this stage, but overall forced alignment was not corrected to the same degree of precision as in the smaller corpus. In the smaller corpus, boundaries at the beginning and end of each stretch were hand-corrected to exclude noise from adjacent obstruents, for example, but in the larger corpus I instead took the approach of includ-

ing the surrounding linguistic context in linguistic models in more detail to account for this. In general, the hand-checking focused more on *excluding* errors rather than correcting them.

Further speakers were excluded at this stage due to poor recording quality or forced alignment problems. As shown in Table 9.1, this resulted in a total of 95 speakers: 19 speakers (9 OF, 3 OM, 3 YF, 4 YM) of Insular Scots, spoken in Orkney and Shetland, 28 speakers from Lothian (7 OF, 7 OM, 7 YF, 7 YM), 48 speakers from Glasgow (10 OF, 10 OM, 13 OF, 15 OM). The varying number of speakers between areas reflect differences in the total available number of each area, as well as differences in usable recordings. This resulted in fewer speakers of Insular Scots included, and more older female speakers than other groups among the Insular Scots speakers.

Area	Older female	Older male	Younger female	Younger male	Total
Glasgow	10	10	13	15	48
Insular	9	3	3	4	19
Lothian	7	7	7	7	28
Total	26	20	23	26	95

Table 9.1: Demographics of speakers in the larger SCOSYA subcorpus

9.2.2 Identification of original voiced stretches

To enable comparison with the smaller corpus analysis, which worked at the level of voiced stretches (stretches of voiced sounds containing little segmental noise, in practice usually sonorants), I again worked at the level of voiced stretches. Here, voiced stretches were defined as stretches of sonorants. Stretches of sonorants were identified for analysis using the output of the forced aligner. 52,684 stretches were originally identified. For the purposes of this analysis, all tokens of /r/ were included, despite the fact that these can also be realised as taps or other non-sonorant sounds in Scottish accents. Voiced stretches in the smaller corpus occasionally included sounds realised as sonorants, even if they were not phonemically so (e.g. /z/ being produced as /z̬/), but here that was not the case: stretches were restricted to phonemic sonorants.

9.2.3 F0-based automatic identification of creak

I followed the same procedure for identifying creak as in the smaller, 24-speaker sample, described in Section 7.3.6. To recap, this follows Dorreen (2017) and Dallaston & Docherty (2019) in using an automatic procedure to code creaky voice when this drops below a speaker-specific f0 threshold, the antimode between a speaker's two f0 modes, the lower of which is assumed to be the mode of their creaky speech and the higher of which is assumed to be the mode of their non-creaky speech. Files were processed at 16

kHz using MacReaper (Dallaston & Docherty 2019), a drag-and-drop implementation of REAPER (Talkin 2015) for Mac OS, to obtain Glottal Closure Instants (GCIs).

The automated procedure considered only GCIs that occurred within stretches of sonorants, as identified above. 52,684 stretches were originally identified, which included 1,188,908 GCIs. After excluding stretches due to local background noise, this left 30,792 stretches containing 613,151 GCIs.

Following Dallaston & Docherty 2019, local f0 was calculated for each GCI by taking the inverse of the time between each GCI within a voiced stretch. Antimodes were then detected using an automated procedure Dallaston & Docherty 2019 in R Core Team 2020, which uses `modes`Deevi & Strategies 2016(v.0.7.0) to identify the f0 mode of non-creaky speech, the f0 mode of creaky speech, and the antimode between them. The version used here is the new version described in Section 7.3.6. Example output is shown in Figure 9.2.

Antimodes were inspected visually for verification. Three speakers were excluded at this stage on the basis that reliable antimodes could not be identified (1 YF Glasgow, 1 OF Glasgow, 1 OF Lothian). This left 92 speakers for analysis.

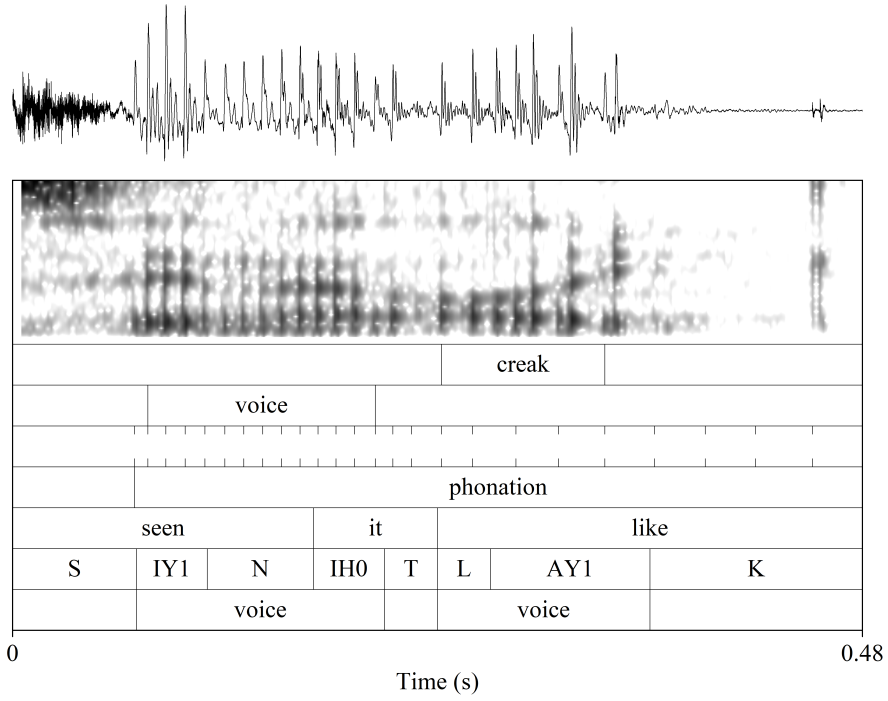
Creaky and non-creaky stretches were separated based on whether local f0 fell below an individual speaker's antimode. In some cases, such as the example shown in Figure 9.1a, this resulted in entire voiced stretches being coded as creaky or not based on f0. However, in other cases, such as in Figure 9.1b, voiced stretches were separated into multiple creaky and non-creaky chunks.

I quantified the percentage of creak used by each speaker and group by dividing the total duration of creaky stretches by the total duration of all sonorant stretches.

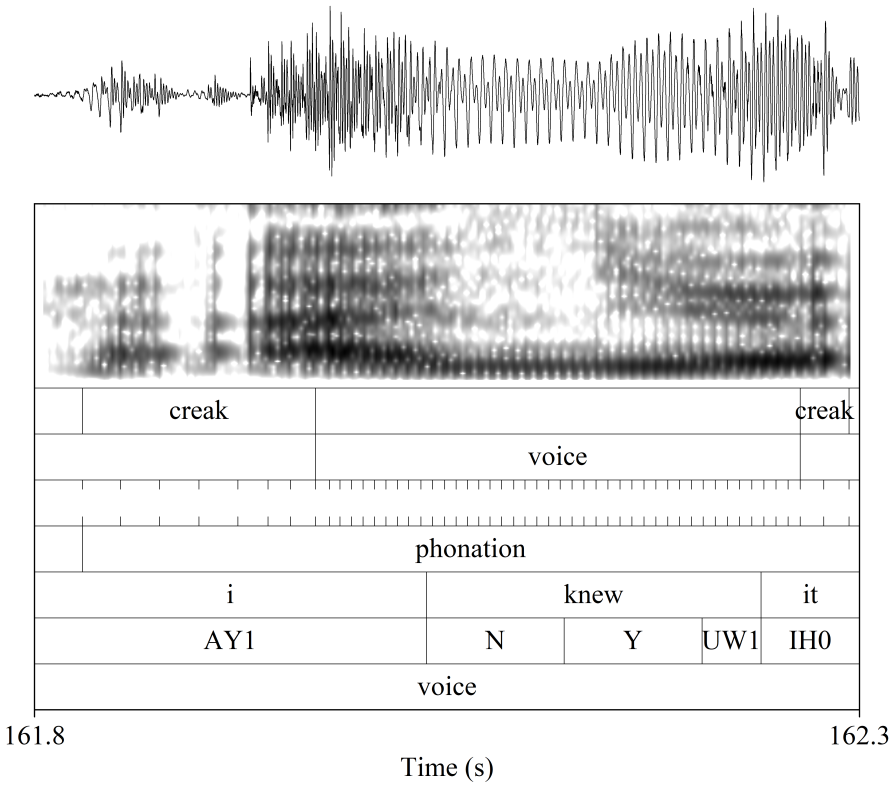
In total, 31,698 stretches were analysed, of which 26922 (85%) were not creaky and 4,776 (15%) were creaky. This was equivalent to a total amount of 3,759s (63 mins) of relevant sonorant stretches, of which 340s (9%) was creaky.

9.2.4 Analysis of non-creaky stretches with VoiceSauce

All 95 speakers were included in this analysis. Where antimodes were identified, stretches with f0 below the speaker antimode were excluded, but for the three speakers who did not have antimodes, all stretches were considered. Acoustic measurements were taken from 16 kHz wav files with VoiceSauce (Shue et al. 2011) using the same procedure as in the smaller corpus, described in Section 8.2. Relevant to subsequent analysis was the measurement of H1*-H2*, H2*-H4*, H4*-2kHz*, and CPP.



(a) YM Glasgow Scott



(b) YF Glasgow Alice

Figure 9.1: Re-chunked voiced stretches for two speakers after f0-based creak coding. Tier 1 shows f0-coded creak while Tier 2 shows stretches with f0 above speaker antimode. Tier 3 shows REAPER's GCIs. The bottom tier shows the original voiced stretches based on consecutive sonorants.

9.2.5 Linguistic factors

Unlike in the smaller corpus, where any linguistic factors considered were hand-coded, all linguistic factors considered in this analysis were automatically extracted from the forced aligner output.

The following linguistic factors were considered:

- Duration: The duration of the voiced stretch, in seconds
- Speech Rate: Local measure of syllables per second, defined as the number of syllables contained in the words that make up the stretch, divided by the duration of those words
- Glottalisation: Whether the stretch occurred in the context of potential glottalisation:
    1. No glottalisation
    2. Pre-glottalisation context: The stretch is followed by a /p/, /t/, /k/ or /tʃ/ that occurs at the end of a syllable, and is followed by a pause or consonant in the case of /ptk/ (e.g. *th**in**k [pause], th**in**k that*) or any sound in the case of /tʃ/ (e.g. *c**a**tching*)
    3. Glottalable context: The stretch is preceded and/or followed by a /t/, which itself occurs at the end of a syllable; is followed by a pause, vowel, /n/, /m/ or /l/; is preceded by a vowel, /l/, /n/ or /r/ (e.g. *b**e**t, b**e**tt**ing**, b**u**tt**on**, b**el**t, b**ent***
- Aspiration: Whether the stretch is preceded by a context that favours aspiration:
    1. No aspiration (e.g. *spine*)
    2. Preceding voiceless plosive aspiration (e.g. *p**ine***)
    3. Preceding voiceless glottal fricative aspiration (e.g. *h**ave***)
- Potential pre-aspiration: Whether the stretch is followed by a voiceless fricative that occurs within the same syllable
    1. No pre-aspiration
    2. Potential pre-aspiration (e.g. *gr**a**ss*)
- Phrase-final position: Whether stretch is followed by a pause identified by the aligner or not
    1. Non-final
    2. Final

- Contains vowel, with the following levels:
    1. Contains a medial vowel
    2. Contains an initial vowel
    3. Contains both a medial and initial vowel
    4. No vowel in stretch
- Contains nasal: Whether the stretch contains /n/,/ŋ/ or /m/, with the levels:
    1. No nasal
    2. Nasal
- Contains rhotic: Whether the stretch contains /r/, with the levels:
    1. No rhotic
    2. Rhotic
- Contains /l/, /w/, or /j/, with the levels:
    1. No /l/, /w/, or /j/
    2. /l/, /w/, or /j/

These linguistic factors were all estimated from the forced aligner output so are best understood as proxies for linguistic factors favouring shifts in phonation; they are, in essence, phonological factors rather than phonetic ones. These factors are not exhaustive: Syllable stress, for example, could well have an impact on phonation but was not included because it was difficult to separate from vowel effects. This is because stress in the MFA output is encoded at the level of the vowel, so it is only possible to extract syllable stress from vowel phones, which would be crossed with other factors, such as whether the stretch contains a vowel. Further, many stretches contained only a single syllable, while others contained many, and it was difficult to compare stress between these contexts.

9.2.6 Statistical analysis

9.2.6.1 Analysis of F0-based creak

The effect of social and linguistic factors on creak was modelled using mixed-effects logistic regression using `glmer` in `lme4` (v1.1-28) (Bates et al. 2015) in R v4.1.2(R Core Team 2020). Mixed-effect logistic regression uses log-transformed odds to predict the probability of a stretch being creaky or not-creaky, as defined by the f0-based method. It allows for the inclusion of random intercepts and slopes that control for

grouping structures in the data (Sonderegger 2021: Ch. 9). For example, modelling an experiment where participants respond to stimuli might benefit from a random intercept for the stimulus id, as participant responses might be grouped by stimulus, but this is unlikely to be the focus of the experiment.

Here, two possible random intercepts were considered:

- `Words`, the orthographic string of words contained in the stretch, with 6225 groups
- `Participant`, with 92 groups, one for each participant

Fixed-effect factors considered for inclusion in the model were:

- `Duration` (Continuous, log-transformed, scaled)
- `Speech Rate` (Continuous, log-transformed, scaled)
- `Glottalisation` (**None**, pre-glottalisation context, glottalable context)
- `Final` (**Non-final**, final)
- `Vowel` (**Medial vowel**, none, initial vowel, both medial and initial)
- `Nasal` (**No nasal**, contains a nasal)
- `Rhotic` (**No rhotic**, contains a rhotic)
- `Contains /l,w,j/` (**No /l/, /w/ or /j/**, contains /l/, /w/ or /j/)
- `Area` (**Glasgow**, Lothian, Insular)
- `Gender` (**Male**, female)
- `Age` (**Older, younger**)

Reference levels for factor variables are shown above in bold. The model was stepped up manually starting from a minimal model containing only the random intercept for `Participant`, with factors added one by one in the order laid out above, and random slopes considered for duration and speech rate. Models were compared in a series of log-likelihood ratio tests, where variables that did not significantly improve the fit of the model were not included. Where the addition of a subsequent variable being added appeared to undo the effect of variable that had already been added to the model, another log-likelihood ratio test was performed with the variable removed.

The final model was as follows:

1. `Creak` $\sim$ `Duration + Speech Rate + Glottalisation + Final + Vowel + Contains /lwj/ + Area + (1 + Duration + Speech Rate | Participant) + (1 + Duration + Speech Rate | Words)`

9.2.6.2 Analysis of acoustic measures

The effect of social and linguistic factors on H1*-H2*, H2*-H4*, H4*-H2kHz* and CPP was modelled using mixed-effects linear regression using `lmer` in `lme4` (v1.1-28) (Bates et al. 2015) in R v4.1.2(R Core Team 2020). Mixed-effect linear regression models the effect of a one-unit change in an independent variable on the outcome variable. It allows for the inclusion of random intercepts and slopes that control for grouping structures in the data.

Here, two possible random intercepts were considered:

- `Words`, the orthographic string of words contained in the stretch, with 6225 groups
- `Participant`, with 92 groups, one for each participant

Fixed-effect factors considered for inclusion in the models were:

- `Duration` (Continuous, log-transformed, scaled)
- `Speech Rate` (Continuous, log-transformed, scaled)
- `Glottalisation` (**None**, pre-glottalisation context, glottalable context)
- `Aspiration` (**None**, aspiration from preceding aspirated /p,t,k/, aspiration from preceding /h/)
- `Preaspiration` (**None**, potential pre-aspiration context)
- `Final` (**Non-final**, final)
- `Vowel` (**Medial vowel**, none, initial vowel, both medial and initial)
- `Nasal` (**No nasal**, contains a nasal)
- `Rhotic` (**No rhotic**, contains a rhotic)
- `Contains /l,w,j/` (**No /l/, /w/ or /j/**, contains /l/, /w/ or /j/)
- `Area` (**Glasgow**, Lothian, Insular)
- `Gender` (**Male**, female)
- `Age` (**Older, younger**)

Reference levels for factor variables are shown above in bold. The model was stepped up manually starting from a minimal model containing only the random intercept for `Participant`, with factors added one by one in the order laid out above, and random slopes considered for duration and speech rate. Models were compared in a series of log-likelihood ratio tests, where variables that did not significantly improve the fit of the model were not included. Where the addition of a subsequent variable being added appeared to undo the effect of variable that had already been added to the model, another log-likelihood ratio test was performed with the variable removed.

The final models were as follows:

1. `H1*-H2* ~ Duration + Speech Rate + Glottalisation + Aspiration + Preaspiration + Vowel + Contains nasal + Contains /l,w,j/ + Gender + Age + (1 + Duration || Participant) + (1 | Words)`

2. `H2*-H4* ~ Glottalisation + Aspiration + Preaspiration + Final + Vowel + Contains nasal + Contains rhotic + Contains /l,w,j/ + Gender + Age + (1 | Participant) + (1 | Words)`

3. `H4*-2kHz* ~ Duration + Glottalisation + Vowel + Contains /l,w,j/ + Age + (1 + Duration || Participant) + (1 + Duration || Words)`

4. `CPP ~ Duration + Aspiration + Final + Vowel + Contains Nasal + Contains Rhotic + Contains /l,w,j/ + Age + (1 + Duration || Participant) + (1 + Duration || words)`

9.3 Results: F0-based creak by social and linguistic factors in larger corpus

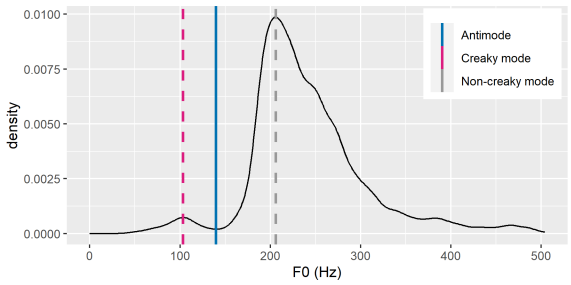
9.3.1 F0-based identification of creaky voice

9.3.1.1 Speaker modes and antimodes

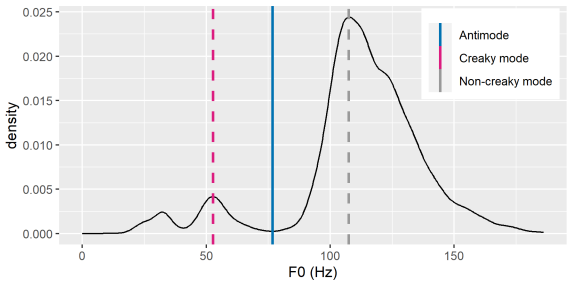
Antimodes were checked visually to ensure they were reliable. Of 95 speakers, reliable antimodes were identified for 92 speakers.

For some speakers, inspection of the f0 distribution suggests that f0 tracking appears to have occurred without major problems, with clear separation between creaky and non-creaky voicing via the antimode. This is the case for Alice, an younger female speaker from Glasgow, whose F0 distribution is shown in Figure 9.2a. However, others show evidence of some minor issues. For example, in the distribution of Aaron (YM Glasgow), shown in Figure 9.2b, two modes can be seen below the antimode, suggesting possible f0 halving. Meanwhile, the distribution of Eleanor (YF Glasgow), shown in Figure 9.2c, is fairly unimodal. However, in both these cases, the antimode identified still appears reasonable.

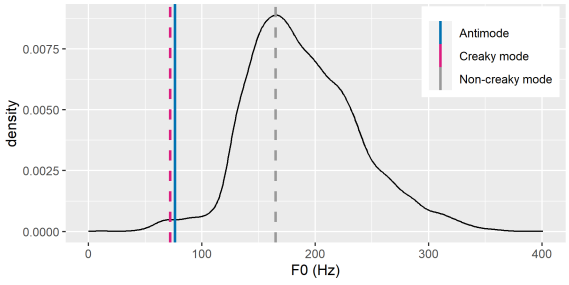
Figure 9.3 shows the unimodal f0 distribution of speakers for whom the identified antimode appears too low when inspected visually. These speakers were excluded from subsequent analysis.



(a) Glasgow YF Alice

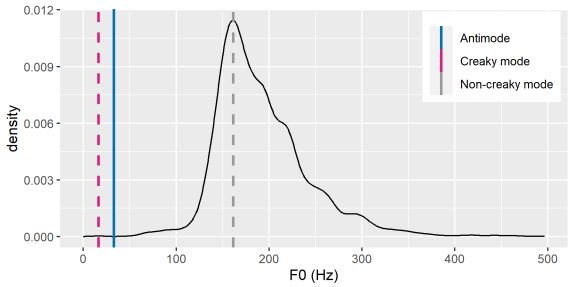


(b) Glasgow YM Aaron

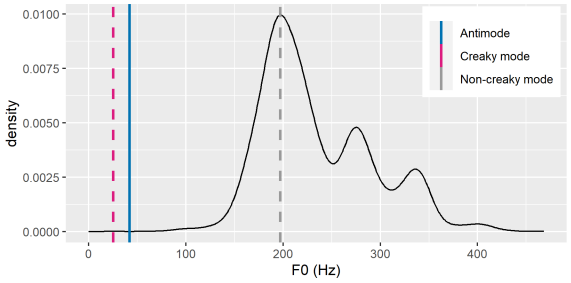


(c) Glasgow YF Eleanor

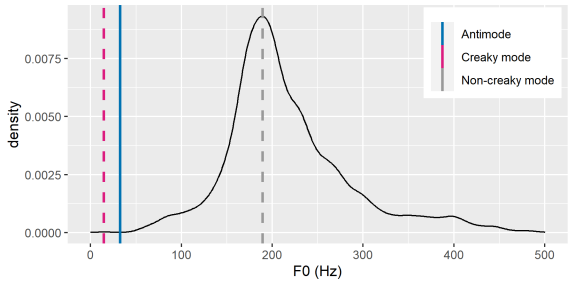
Figure 9.2: F0 distributions with modes and antimodes identified for three speakers



(a) Glasgow OF Sandra



(b) Glasgow YF Siobhan



(c) Lothian OF Margaret

Figure 9.3: F0 distributions with poor antimode identification for three speakers

9.3.1.2 Percentage of creak by speaker and social factors

Speakers used creak 9.0% of the time in speech considered in this analysis. There do not appear to be major deviations from this in terms of gender when these are considered across all areas, with male speakers using creak 9.0% of the time and female speakers using creak 9.1% of the time. Younger speakers use creak 9.9% of the time, compared to 8.1% of the time for older speakers. Insular speakers appear particularly creaky, using creak 13.2% of the time compared to Glasgow speakers' 8.2% of the time and Lothian speakers' 6.7%.

However, as summarised in Table 9.2 and shown in Figure 9.4, the percentage of creak used by speakers out of all voiced speech differs in raw numeric terms when differences between speakers of Scots varieties and different age groups are considered. Older male Insular speakers, for example, use creak 18.7% of the time, while older male Lothian speakers use creak 4.8% of the time.

9.3.1.3 Statistical analysis

9.3.1.3.1 Model diagnostics Scaled residuals were simulated from the model and plotted using the `DHARMa` package (Hartig 2022) to judge them against a uniform distribution. As shown in Figure 9.5, the scaled residuals do not deviate significantly from uniformity, which is confirmed through Kolmogorov-Smirnov test (noted as KS test on the figure).

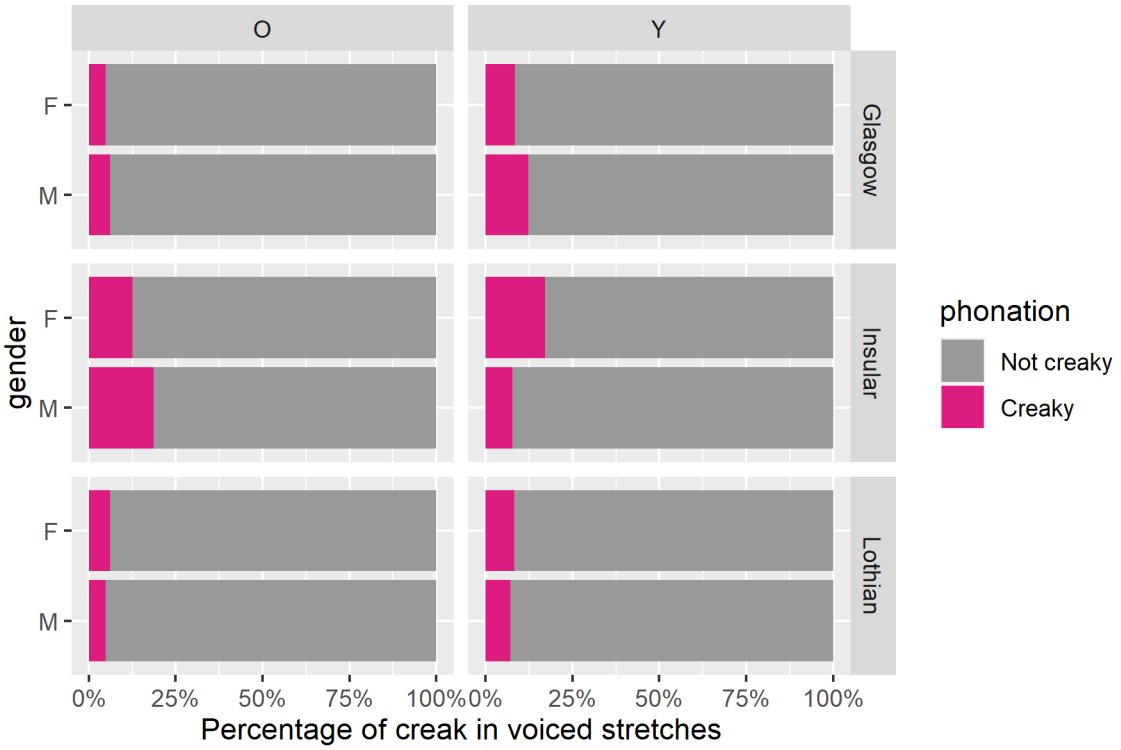


Figure 9.4: Percentage of time creak used in voiced stretches by area, age and gender

There is no evidence of problematic collinearity in the model with all GVIFs < 1.5 and $c = 1.5$ for linear predictors.

Table 9.5 shows pseudo-R^2 measures for the full and baseline model. This suggests more variance is explained by the addition of fixed effects and an additional random intercept for Words.

Figure 9.6 shows the effects of linguistic factors on the predicted probability of a stretch being creaky. Increased voiced stretch Duration and Speech Rate both reduced the log-odds of creak.

By contrast, creak was favoured by Phrase Final position, and a stretch occurring in a Glottalable Context. A stretch being followed by a potential Pre-Glottalisation context did not favour creak. Furthermore, containing an Initial Vowel and containing No Vowel also favoured creak, while containing both a Medial and Initial Vowel did not favour creak. Containing /l,w,j/ also favoured creak compared to not containing /l,w,j/.

Insular Scots voice quality stands out as particularly creaky: Being an Insular speaker increased the log(odds) of a stretch being creaky. The effect of Area on the predicted probability of a stretch being creaky is shown in Figure 9.7

The effect of random effects is shown in Table 9.4.

Area	Age	Gender	Creak (s)	Voice (s)	Total duration (s)	% creak
Glasgow	O	M	24.4	376.5	400.8	6.1
		F	18.4	362.7	381.0	4.8
	Y	M	62.3	443.2	505.5	12.3
		F	40.0	434.0	474.0	8.4
Insular	O	M	22.3	96.8	119.2	18.7
		F	56.3	394.4	450.7	12.5
	Y	M	14.9	178.4	193.4	7.7
		F	30.2	146.4	176.6	17.1
Lothian	O	M	11.1	221.4	232.5	4.8
		F	13.9	211.8	225.7	6.1
	Y	M	21.0	270.9	291.9	7.2
		F	25.1	282.2	307.4	8.2
Total	-	-	339.9	3418.7	3758.7	9.0

Table 9.2: Percentage of time creak used in voiced stretches by area, age and gender

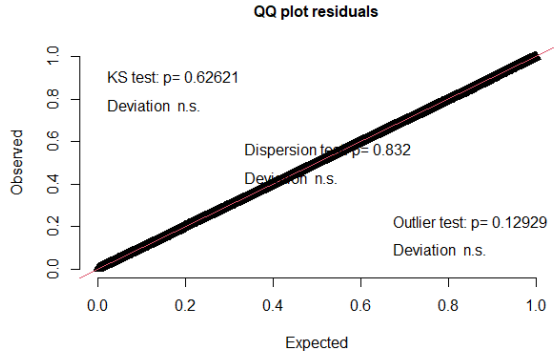


Figure 9.5: Scaled (quantile) residuals for the mixed-effects logistic regression predicting creak

9.3.2 Acoustic measures

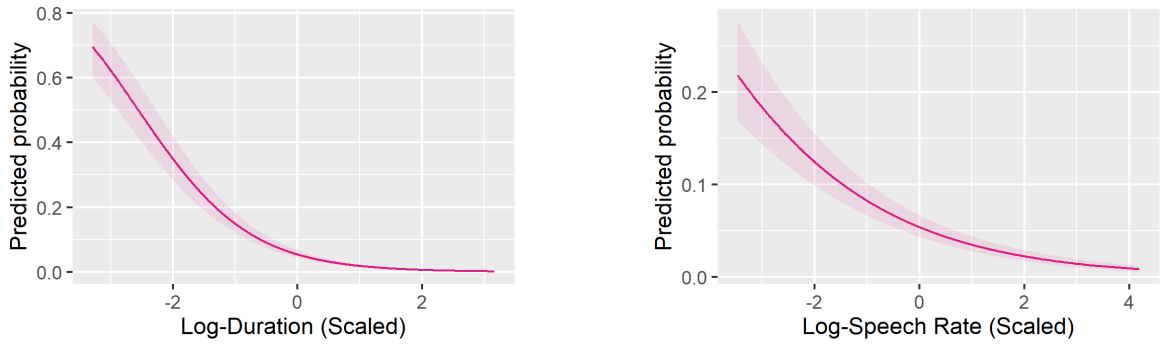
9.3.2.1 Model diagnostics

Figure 9.8 shows QQ plots of the residuals of each model. The residuals for CPP are close to normality, while the residuals for each of the spectral slope measures show some deviation from normality. However, as Sonderegger (2021: 102) explains, non-normal distribution of residuals is a common violation of regression assumptions in larger samples, and is acceptable where the goal of regression is to estimate coefficients rather than predict new data.

There is no evidence of problematic collinearity in any model with all GVIFs < 1.5 and $c = 1.5$ for linear predictors.

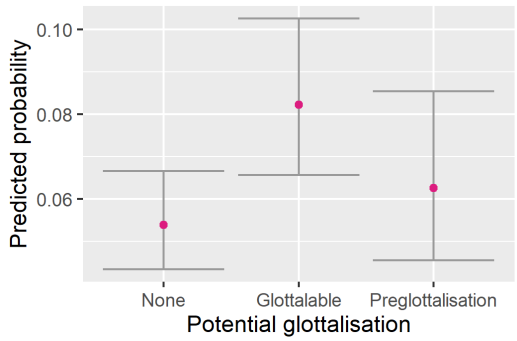
Table 9.3: Results of the mixed-effect logistic regression model predicting the log-odds of a stretch being coded as creaky as a function of Duration, Speech Rate, Glottalisation, Phrase Position, Vowel, Contains /l,w,j/ and Area

Independent variable	Level/unit	Coefficient (SE)
Intercept		−2.867 (0.116)
		t = −24.783***
Duration	Scaled log-duration	−1.121 (0.054)
		t = −20.724***
Speech Rate	Scaled log-Speech Rate	−0.458 (0.033)
		t = −13.708***
Glottalisation	Glottalable context	0.454 (0.057)
(Ref = No Glottalisation)		t = 7.943***
	Pre-Glottalisation context	0.159 (0.135)
		t = 1.184
Phrase position	Final	0.475 (0.045)
(Ref = Not Final)		t = 10.471***
Vowel	Both Initial and Non-Initial	−0.076 (0.303)
(Ref = Non-Initial Vowel)		t = −0.252
	Initial Vowel	0.660 (0.085)
		t = 7.806***
	None	1.445 (0.088)
		t = 16.484***
Contains /l,w,j/	Contains /l,w,j/	0.177 (0.052)
(Ref = No /l,w,j/)		t = 3.383***
Area	Insular	0.610 (0.197)
(Ref = Glasgow)		t = 3.101**
	Lothian	−0.189 (0.178)
		t = −1.058
	Log Likelihood	−10,413.780
	Observations	31,697
	Note:	*p<0.05; **p<0.01; ***p<0.001

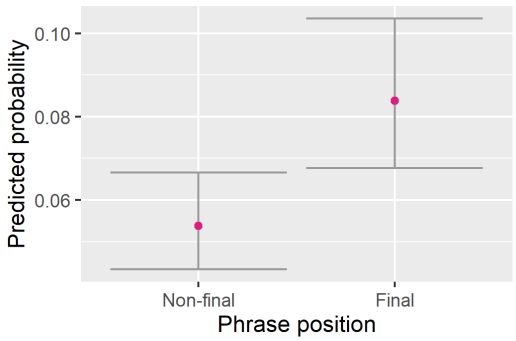


(a) Duration

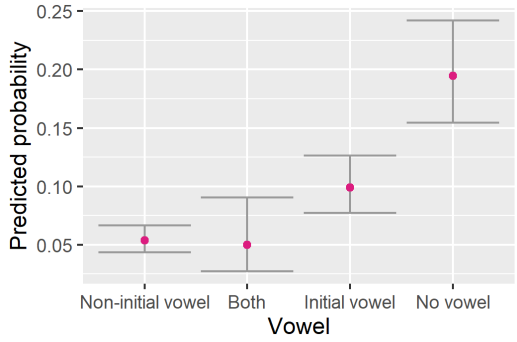
(b) Speech Rate



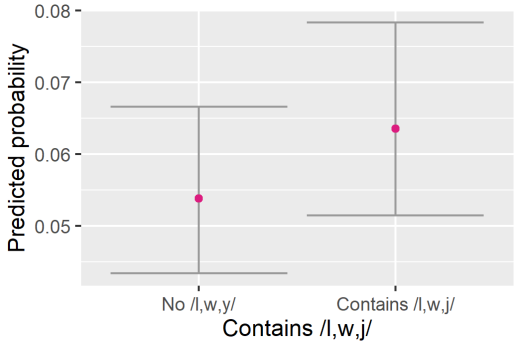
(c) Potential Glottalisation



(d) Phrase Final position



(e) Vowel



(f) Contains /l,w,j/

Figure 9.6: The effect of linguistic factors on predicted probability of a stretch being creaky (all other variables held constant). Note that y-axes are not equivalent between figures.

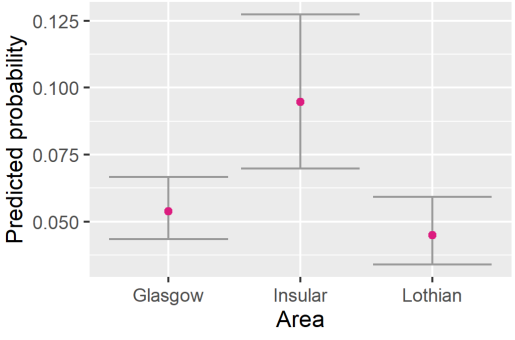
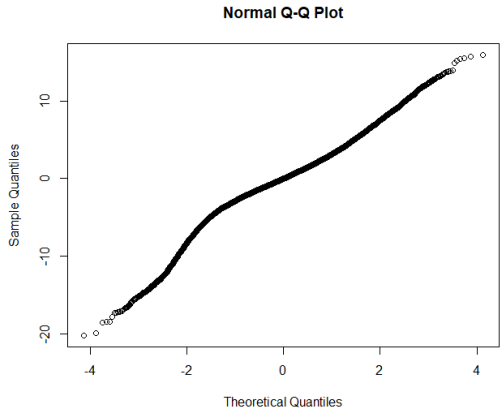
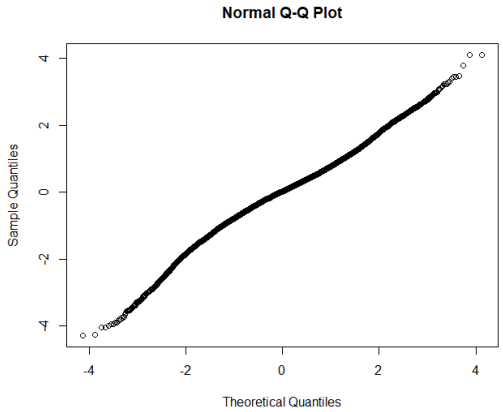


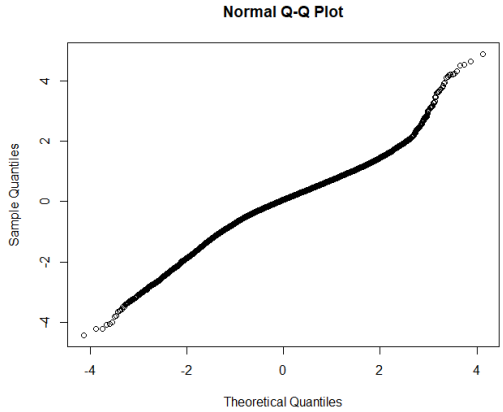
Figure 9.7: The effect of Area on the predicted probability of a stretch being creaky



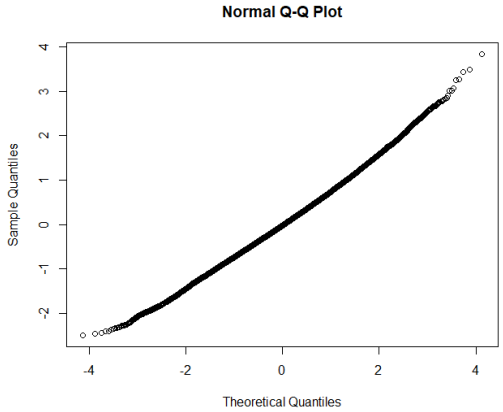
(a) H1*–H2*



(b) H2*–H4*



(c) H4*–2kHz*



(d) CPP

Figure 9.8: QQ plots of the residuals for the models for H1*-H2*, H2*-H4*, H4*-2kHz and CPP.

Groups	Name	Variance	Std. deviation
Words	Speech Rate	0.03164	0.1779
Words.1	Duration	0.19999	0.4472
Words.2	(Intercept)	0.30227	0.5498
Participant	Speech Rate	0.01049	0.1024
Participant.1	Duration	0.15334	0.3916
Participant.2	(Intercept)	0.48103	0.6936

Table 9.4: Effect of random intercepts in the mixed-effects binary logistic regression model for creak

Model		R2m	R2c
Full	theoretical	0.24	0.44
	delta	0.13	0.24
Baseline	theoretical	0	0.15
	delta	0	0.07

Table 9.5: Pseudo R^2 measures for baseline and full model. Baseline model includes only a random intercept for participant.

9.3.2.2 H1*–H2*

Table 9.6 shows the results of the mixed-effects linear regression predicting H1*–H2*. Random effects are presented in Table 9.7, and R^2 is given in Table 9.8.

To recap, H1*–H2* is a measure of the degree of glottal constriction, and in the smaller corpus analysis, I found that breathy stretches showed higher H1*–H2* than whispery stretches, while tense stretches showed lower H1*–H2* than whispery stretches. I therefore interpret H1*–H2* in terms of a breathy-tense continuum, and in Figure 9.9 and Figure 9.10 which visualise the results of the model presented in Table 9.6, I use purple to represent breathier values, and pink to represent tenser values.

Figure 9.9 shows the effect of linguistic factors on H1*–H2*.

As shown in Figure 9.9a, as stretch Duration increased, predicted H1*–H2* decreased, indicating that longer stretches tend to be produced with tenser phonation.

Furthermore, as shown in Figure 9.9b, a stretch occurring in a Glottalable Context (i.e. preceding or following a /t/ that can be produced as a glottal stop) significantly decreased predicted H1*–H2*, indicating tenser phonation in Glottalable Contexts. There was no effect of Pre-Glottalisation.

As shown in Figure 9.9c, aspiration from a preceding /h/ significantly increased predicted H1*–H2*, indicating breathier phonation in voiced stretches following /h/. However aspiration from a preceding voiceless plosive had no significant effect.

Preaspiration also significantly increased predicted H1*–H2*. This is shown in Figure 9.9d, which shows how Pre-Aspiration contexts are produced with breathier

Table 9.6: Results of the mixed-effects linear regression predicting H1*–H2* as a function of social and linguistic factors.

Independent variable	Level/unit	Coefficient (SE)
Intercept		0.362 (0.078) t = 4.660***
Duration	Scaled log-duration	−0.021 (0.009) t = −2.241*
Glottalisation	Glottalable context	−0.052 (0.018) t = −2.878**
	Pre-Glottalisation context	−0.066 (0.039) t = −1.695
Aspiration	Aspiration from preceding /p,t,k/	0.018 (0.015) t = 1.216
	Aspiration from preceding /h/	0.105 (0.030) t = 3.516***
Preaspiration	Pre-aspiration	0.079 (0.024) t = 3.343***
Vowel	Both Initial and Non-Initial	0.075 (0.050) t = 1.514
	Initial vowel	0.065 (0.029) t = 2.256*
	None	−0.280 (0.047) t = −5.935***
Contains Nasal	Contains Nasal	−0.147 (0.015) t = −9.572***
Contains /l,w,j/	Contains /l,w,j/	0.032 (0.014) t = 2.256*
Gender	F	−0.354 (0.084) t = −4.209***
Age	Y	−0.234 (0.084) t = −2.776**
Observations	27,563	
Log Likelihood	−36,404.920	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Groups	**Name**	**Variance**	**Std. Dev.**
Words	(Intercept)	0.012653	0.1125
Participant	Duration	0.003248	0.0570
Participant	(Intercept)	0.163227	0.4040
Residual		0.797182	0.8929

Table 9.7: Effects of the random intercepts and slopes in the model predicting H1*–H2*

Model	R2m	R2c
Full	0.05	0.22
Baseline	0	0.19

Table 9.8: Pseudo R^2 measures for baseline and full model for H1*–H2. Baseline model includes only a random intercept for participant.

phonation.

Compared to stretches containing a Non-Initial Vowel, containing an Initial Vowel significantly increased predicted H1*–H2*. Meanwhile, containing No Vowel significantly decreased H1*–H2*. This effect, shown in Figure 9.9e, suggests that Intital Vowels are produced with breathier phonation, while stretches that do not contain vowels are produced with tenser phonation.

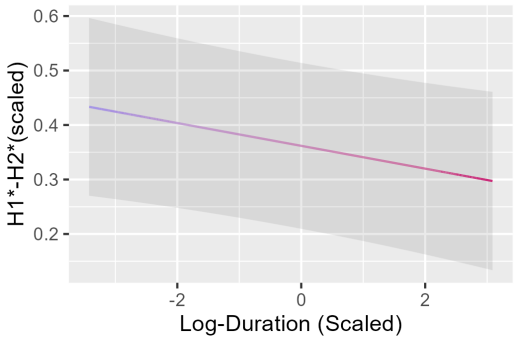
Furthermore, as shown in Figure 9.9f containing a nasal decreased H1*–H2*, suggests containing a nasal favoured tenser phonation. On the other hand, as shown in Figure 9.9g, containing /l,w,j/ increased H1*–H2*, suggesting that containing one of these sounds favoured breathier phonation.

The effects of social factors on H1*–H2* are shown in Figure 9.10. As shown in Figure 9.10a, Female speakers showed significantly lower values for H1*–H2* when compared to Male speakers, indicating that Female speakers use tenser phonation. As shown in Figure 9.10b Younger speakers also showed significantly lower H1*–H2* than Older speakers , indicating that Younger speakers use tenser phonation.

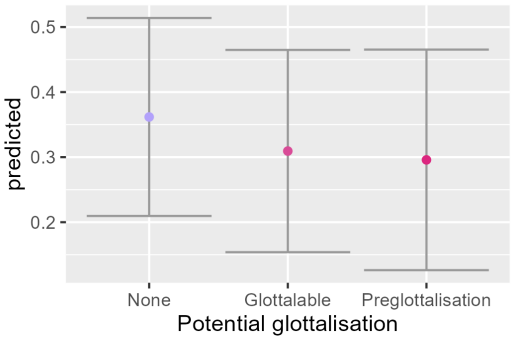
9.3.2.3 H2*–H4*

Table 9.9 summarises the results of the model H2*–H4*. Table 9.10 gives the results of random effects, and Table 9.11 shows R^2.

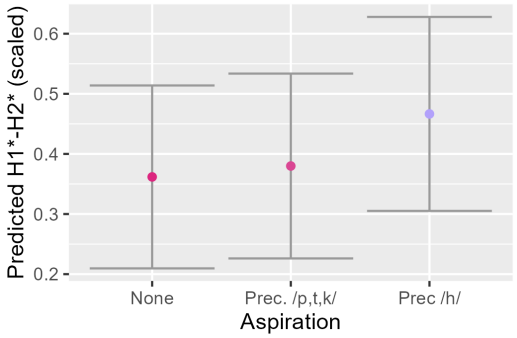
To recap, H2*–H4* is a measure of glottal constriction, and in the smaller corpus analysis presented in Section 8.3, I found that stretches with higher H2*–H4* were more likely to be modal rather than whispery, while progressively higher H2*–H4* related to increasing breathy voice. I therefore interpret H2*–H4* as relating to a whispery-modal-breathy continuum, where lower values indicate a more whispery quality, and higher values suggest a modal-breathy quality. In the plots shown in Figure 9.11



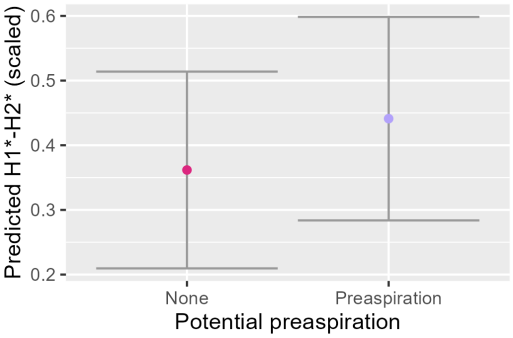
(a) Duration



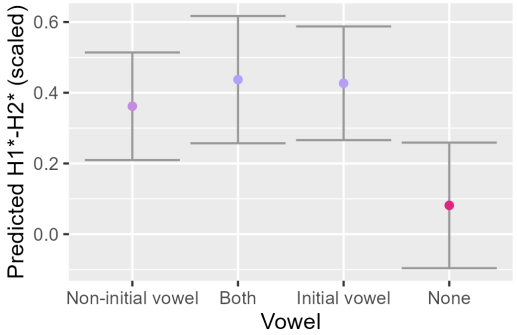
(b) Potential Glottalisation



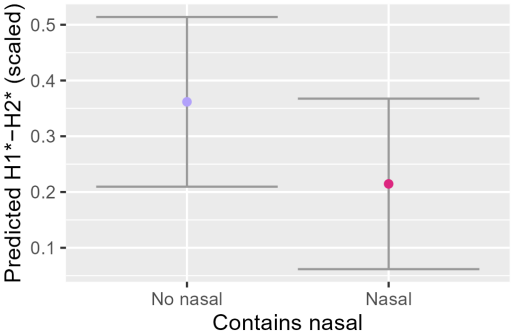
(c) Aspiration



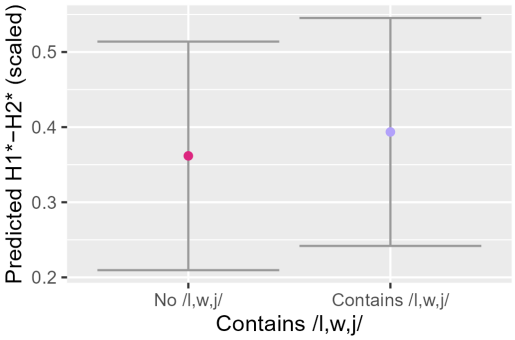
(d) Potential pre-aspiration



(e) Vowel



(f) Nasal

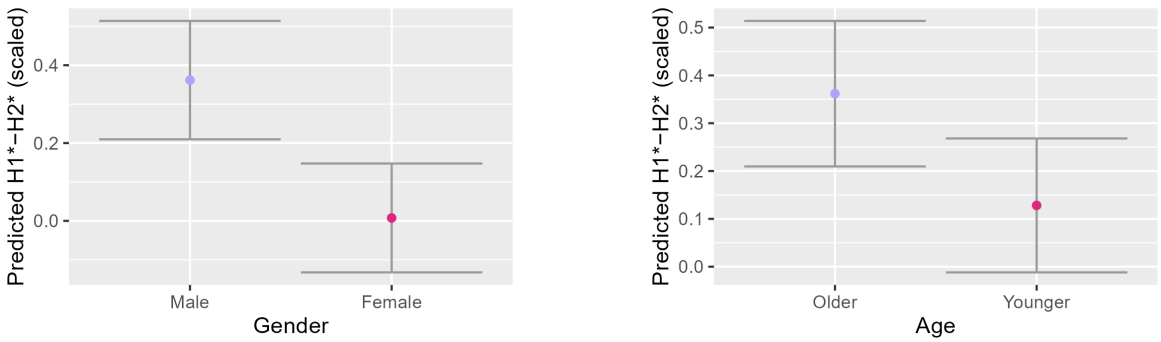


(g) Contains /l,w,j/

Figure 9.9: The effect of linguistic factors on predicted H1*–H2* (scaled) (all other variables held constant). Note that y-axes are not equivalent between figures.

Table 9.9: Results of the mixed-effects linear regression model predicting H2*–H4* as a function of social and linguistic factors

Independent variable	Level/unit	Coefficient (SE)
Intercept		0.511 (0.057) t = 8.907***
Glottalisation	Glottalable context	0.080 (0.019) t = 4.286***
	Pre-Glottalisation context	0.020 (0.040) t = 0.500
Aspiration	Aspiration from preceding /p,t,k/	−0.029 (0.015) t = −1.902
Aspiration	Aspiration from preceding /h/	0.061 (0.033) t = 1.884
Preaspiration	Pre-aspiration	0.102 (0.025) t = 4.139***
Phrase position	Final	0.035 (0.013) t = 2.810**
Vowel	Both Initial and Non-Initial	0.174 (0.050) t = 3.503***
	Initial vowel	0.067 (0.029) t = 2.306*
	None	−0.208 (0.047) t = −4.388***
Contains Nasal	Contains Nasal	−0.038 (0.015) t = −2.523*
Contains Rhotic	Contains Rhotic	0.069 (0.017) t = 4.051***
Contains /l,w,j/	Contains /l,w,j/	−0.164 (0.014) t = −11.382***
Gender	F	−0.696 (0.061) t = −11.451***
Age	Y	−0.157 (0.061) t = −2.586**
Observations	27,563	
Log Likelihood	−35,746.130	
Note:	*p<0.05; **p<0.01; ***p<0.001	



(a) Gender

(b) Gender

Figure 9.10: The effect of social factors on predicted H1*–H2* (scaled) (all other variables held constant). Note that y-axes are not equivalent between figures.

Words	(Intercept)	0.03410	0.1847
Participant	(Intercept)	0.08374	0.2894
Residual	0.75288	0.8677	

Table 9.10: Variance explained by random effects in model predicting H2*–H4*

and Figure 9.12, I therefore use yellow to represent values that are more likely to be whispery, and purple to represent values that are more likely to be breathy.

As shown in Figure 9.11a, occurring in Glottalable Context increased predicted H2*–H4*. This suggests that occurring in a Glottalable context favours modal-breathy phonation over whispery-tense phonation.

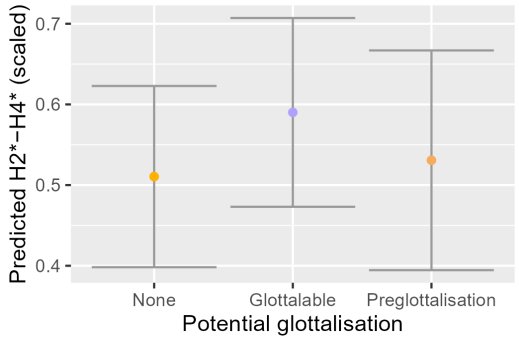
There was no significant effect of preceding aspiration. However, being followed by potential Preaspiration increased predicted H2*–H4*, as shown in Figure 9.11c. This suggests that following potential Preaspiration favours modal-breathy phonation over whispery-tense phonation.

Occurring in Final position also increased predicted H2*–H4*, as shown in Figure 9.11d. This suggests that occuring in Phrase-Final Position favours modal-breathy phonation over whispery-tense phonation.

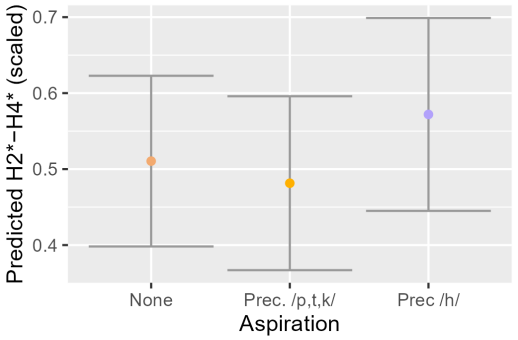
The effect of Vowel is shown in Figure 9.11e. Compared to containing a Non-Initial vowel, containing either an Initial Vowel or Both Initial and Non-Initial vowels increased H2*–H4* = 0.050, t = 3.503, $p < 0.001$). Meanwhile, not containing a vowel decreased predicted H2*–H4*. This suggests that containing an initial vowel (either alongside a non-initial vowel, or alone) favours modal-breathy phonation, while containing no

Model	R2m	R2c
Full	0.16	0.4
Baseline	0	0.21

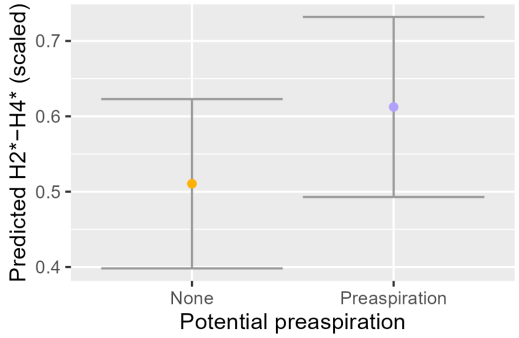
Table 9.11: Pseudo R^2 measures for baseline and full model for H2*–H4*. Baseline model includes only a random intercept for participant.



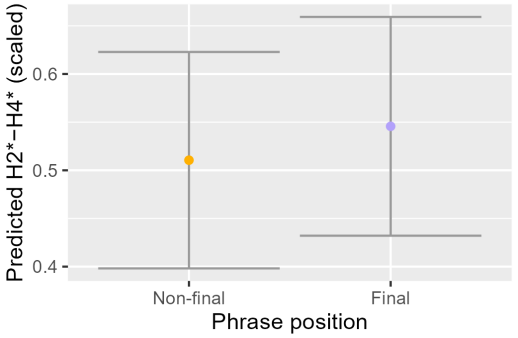
(a) Potential Glottalisation



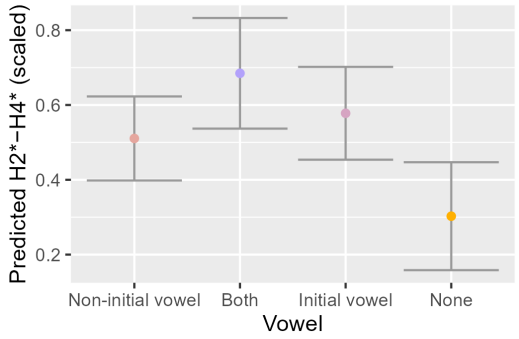
(b) Aspiration



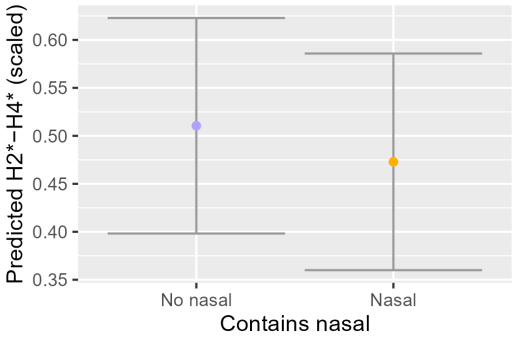
(c) Potential pre-aspiration



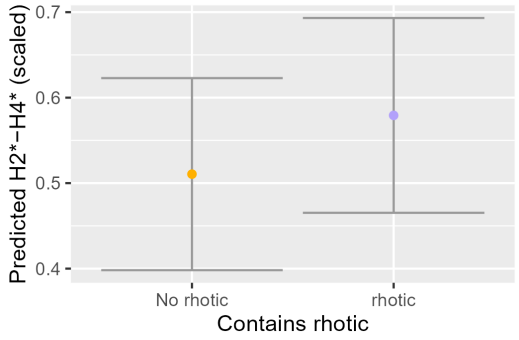
(d) Phrase position



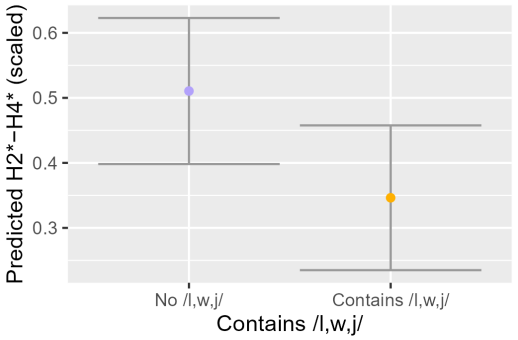
(e) Vowel



(f) Nasal



(g) Contains Rhotic



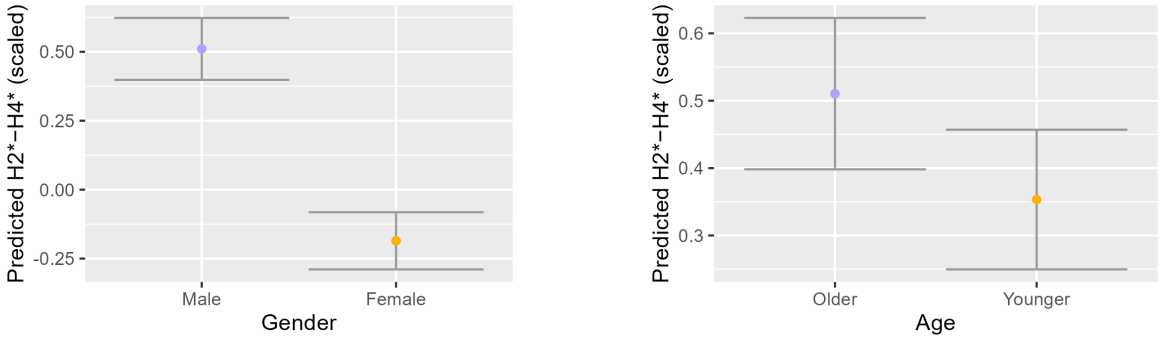
(h) Contains /l,w,j/

Figure 9.11: The effect of linguistic factors on predicted H2*–H4* (scaled) (all other variables held constant). Note that y-axes are not equivalent between figures.

vowel favours whispery-tense phonation.

The presence of a Nasal in a stretch decreased predicted H2*–H4*, while containing a rhotic increased predicted H2*–H4*, and containing /l,w,j/ decreased predicted H2*–H4*. These effects are shown in Figures 9.11f, 9.11g and 9.11h. They indicate that the presence of nasals and /l,w,j/ favour whispery-tense phonation, while rhotics favour modal-breathy phonation.

The effects of social factors are shown in Figure 9.12. Female speakers had significantly lower H2*–H4* than Male speakers, while Younger speakers had significantly lower H2*–H4* when compared to Older speakers. This suggests that female speakers and younger speakers favour whispery-tense phonation, while male speakers and older speakers favour modal-breathy phonation.



(a) Gender (b) Age

Figure 9.12: The effect of social factors on predicted H2*–H4* (scaled) (all other variables held constant). Note that y-axes are not equivalent between figures.

9.3.2.4 H4*–2kHz*

Table 9.12 shows the results mixed-effects linear regression model predicting H4*–2kHz* as a function of social and linguistic factors. Random effects are given in Table 9.14 and R^2 is given in Table 9.13.

To recap, H4*–2kHz* is a measure of spectral tilt in a higher-frequency region of the spectrum, where higher values suggest the presence of a glottal gap. Based on the finding presented in Chapter 8.3, I interpret H4*–2kHz* here as a measure of whisperiness, where lower values suggest a more whispery quality, and higher values suggest a more breathy or more tense whispery quality. In the plots shown in Figure 9.13, I therefore use yellow to represent values that are more likely to be whispery, and pink-orange to represent values that are more likely to be tense whispery.

As shown in Figure 9.13a, as the Duration of a stretch increased, H4*–2kHz* decreased. This suggests that longer stretches are produced with less of posterior glottal gap, and are less likely to be breathy or tense whispery.

Table 9.12: Results of the mixed-effects linear regression model predicting H4*–2kHz* as a function of social and linguistic factors

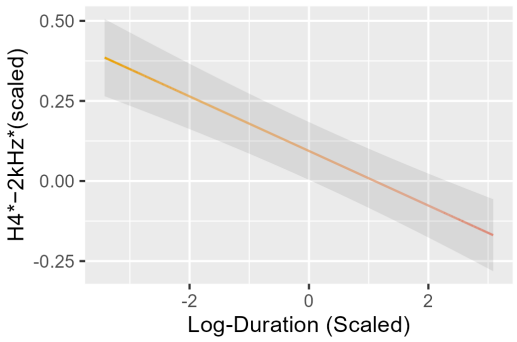
Independent variable	Level/unit	Coefficient (SE)
Intercept		0.094 (0.046) t = 2.040*
Duration	Scaled log-duration	−0.085 (0.012) t = −7.386***
Glottalisation	Glottalable context	−0.063 (0.019) t = −3.334***
	Pre-Glottalisation context	−0.111 (0.046) t = −2.423*
Vowel	Both Initial and Non-Initial	−0.100 (0.054) t = −1.848
	Initial vowel	−0.077 (0.029) t = −2.627**
	None	−0.430 (0.051) t = −8.492***
Contains /l,w,j/	Contains /l,w,j/	−0.056 (0.018) t = −3.172**
Age	Y	−0.149 (0.061) t = −2.447*
Observations	27,563	
Log Likelihood	−35,998.190	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Model	R2m	R2c
Full	0.02	0.31
Baseline	0	0.09

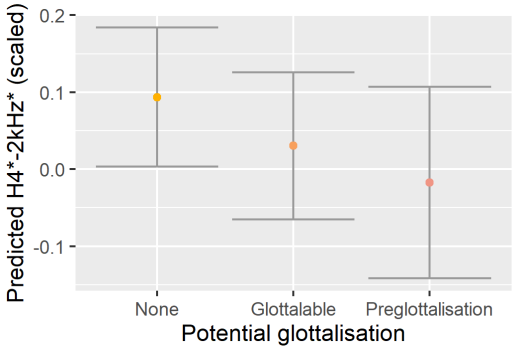
Table 9.13: Pseudo R^2 measures for baseline and full model for H4*–2kHz*. Baseline model includes only a random intercept for participant.

Groups	Name	Variance	Std.Dev.
Words	Duration	0.008706	0.09331
Words	(Intercept)	0.200177	0.44741
Participant	Duration	0.005787	0.07607
Participant	(Intercept)	0.084698	0.29103
Residual		0.698643	0.83585

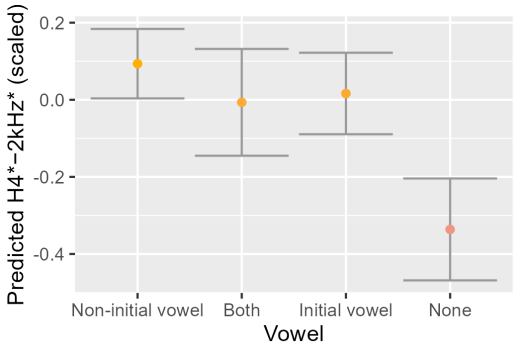
Table 9.14: Effect of random intercepts and slopes in the model predicting H4*–2kHz*



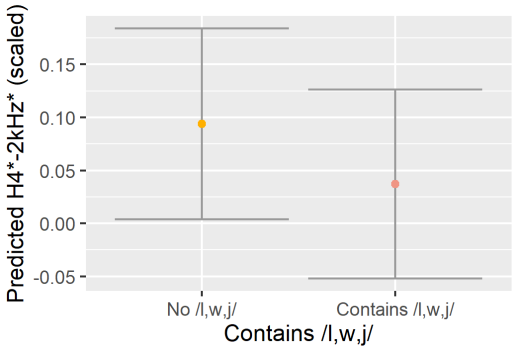
(a) Duration



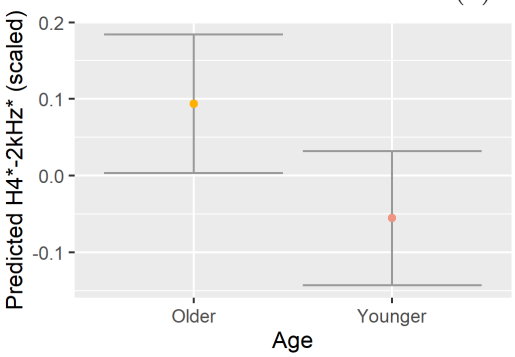
(b) Potential Glottalisation



(c) Vowel



(d) Contains /l,w,j/



(e) Age

Figure 9.13: The effect of each independent variable on predicted H4*–2kHz* (scaled) (all other variables held constant). Note that y-axes are not equivalent between figures.

As shown in Figure 9.13b, occurring in a Glottalable Context decreased H4*–2kHz*, as did occurring in a Pre-Glottalisation Context. This suggests that glottalisation is produced with less of posterior glottal gap, is less likely to be breathy or tense whispery.

Compared to containing a Non-Initial vowel, containing an Initial Vowel decreased H4*–2kHz*, while not containing a vowel decreased H4*–2kHz*. This suggests that stretches containing initial vowels, or lacking vowels, are produced with less of posterior glottal gap, and are less likely to be breathy or tense whispery. This effect is shown in Figure 9.13c.

As shown in Figure 9.13d, containing /l,w,j/ also decreased predicted H4*–2kHz*, compared to not containing /l,w,j/. This suggests that stretches that contained /l,w,j/ are produced with less of posterior glottal gap, and are less likely to be breathy or tense whispery.

Being a Younger speaker also decreased predicted H4*–2kHz*. This suggests that younger speakers use less of a posterior glottal gap, and are less likely to use breathy or tense whispery voice quality.

9.3.2.5 CPP

Table 9.15 shows the results of the model predicting CPP as a function of social and linguistic factors. Random effects are shown in Figure 9.16 and R^2 is given in Table 9.17.

To recap, CPP is a measure of aperiodic noise, where higher values suggest less noise and lower values suggest more noise. In the smaller corpus, modal and tense stretches showed higher CPP, while whispery stretches showed lower CPP. In the plots showing the results of a the model predicting CPP as a function of social and linguistic factors shown in Figure 9.15, I therefore use grey to represent values that are more likely to be near-modal, and yellow to represent stretches that are more likely to be more whispery.

Here, stretches with a longer Duration had a significantly higher CPP, suggesting that longer stretches are less noisy. This is shown in Figure 9.14a.

As shown in Figure 9.14b, Aspiration from preceding /p,t,k/ significantly decreases CPP, suggesting that Aspiration from voiceless plosives increases the amount of aperiodic noise in a stretch. Aspiration from preceding /h/ does not have a significant effect on CPP.

Occurring in Phrase Final position decreases CPP significantly. This suggests that stretches that occur in Phrase Final position are noisier.

Table 9.15: Results of the mixed-effects linear regression model predicting CPP as a function of social and linguistic factors

Independent variable	Level/unit	Coefficient (SE)
Intercept		0.109 (0.061) t = 1.784
Duration	Scaled log-duration	0.440 (0.013) t = 32.680***
Aspiration	Aspiration from preceding /p,t,k/	−0.079 (0.014) t = −5.744***
Aspiration	Aspiration from preceding /h/	0.039 (0.033) t = 1.189
Phrase position	Final	−0.202 (0.012) t = −17.121***
Vowel	Both Initial and Non-Initial	−0.230 (0.047) t = −4.876***
	Initial vowel	−0.287 (0.027) t = −10.697***
	None	−0.496 (0.046) t = −10.817***
Contains Nasal	Contains Nasal	−0.393 (0.016) t = −24.391***
Contains Rhotic	Contains Rhotic	−0.104 (0.017) t = −6.142***
Contains /l,w,j/	Contains /l,w,j/	−0.140 (0.015) t = −9.270***
Age	Y	0.273 (0.083) t = 3.304***
Observations	27,563	
Log Likelihood	−33,151.110	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Groups	Name	Variance	Std.Dev.
Words	Duration	0.008299	0.0911
Words	(Intercept)	0.063593	0.2522
Participant	Duration	0.010502	0.1025
Participant	(Intercept)	0.159274	0.3991
Residual		0.596113	0.7721

Table 9.16: Effect of random intercepts and slopes in the model for CPP

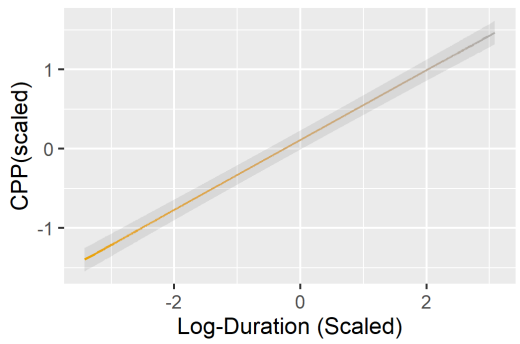
Model	R2m	R2c
Full	0.16	0.4
Baseline	0	0.17

Table 9.17: Pseudo R^2 measures for baseline and full model for CPP. Baseline model includes only a random intercept for participant.

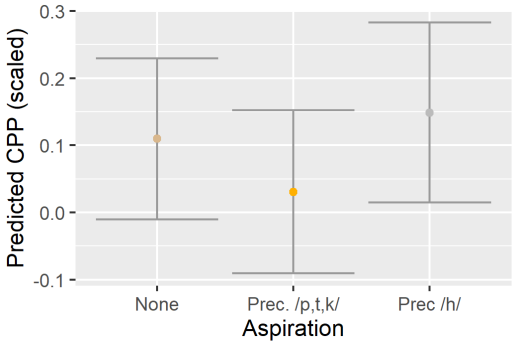
Compared to containing a Non-Initial vowel, containing No Vowel, an Initial vowel, or Both Initial and Non-Initial vowels decreased CPP. This effect, shown in Figure 9.14d suggests that stretches that compared to containing a Non-Initial vowel only, other contexts are noisier.

Containing any sound other than a vowel also significantly decreases CPP. This suggests that consonants involve more aperiodic noise than vowels. This is shown in Figure 9.14e, Figure 9.14f and Figure 9.14g.

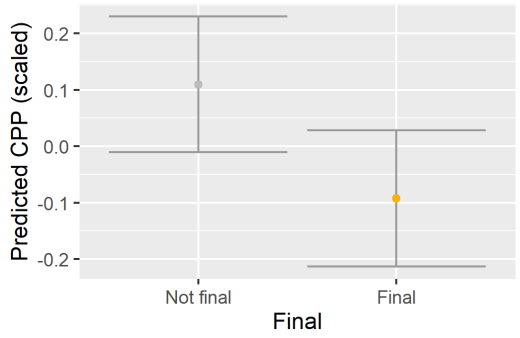
Compared to Older speakers, Younger speakers had a significantly higher CPP. This suggests that younger speakers are less noisy, or more tense-modal, than older speakers.



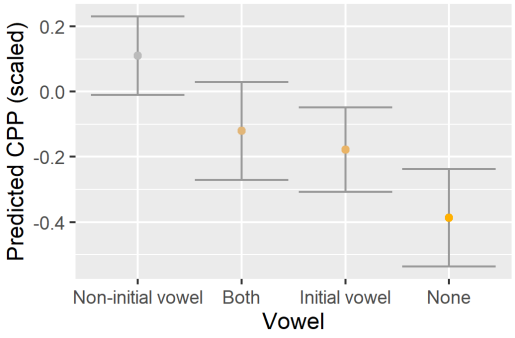
(a) Duration



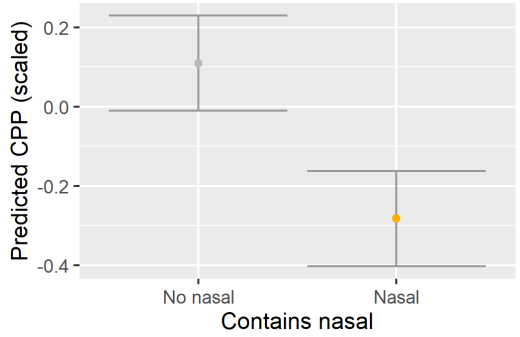
(b) Potential aspiration



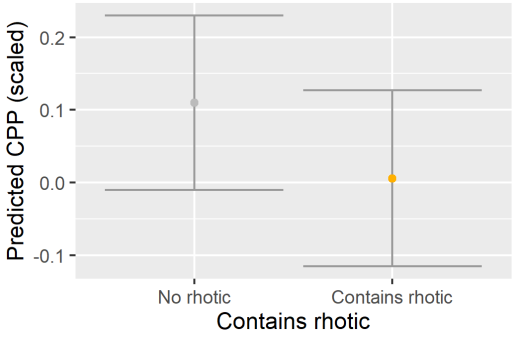
(c) Final



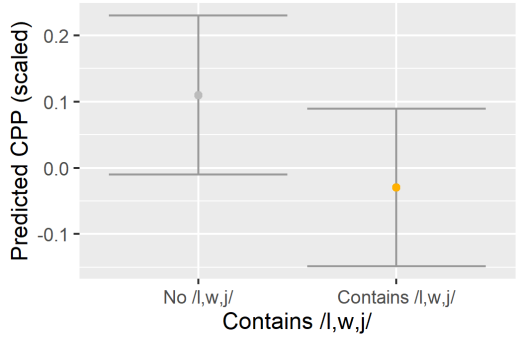
(d) Vowel



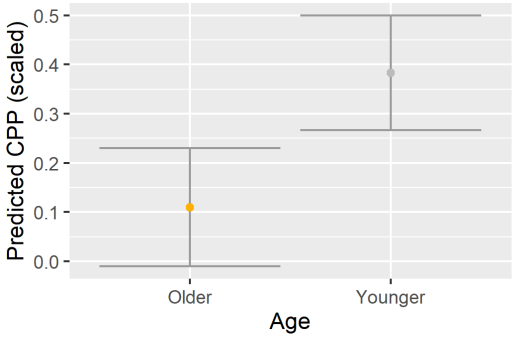
(e) Nasal



(f) Rhotic



(g) Contains /l,w,j/



(h) Age

Figure 9.14: The effect of each independent variable on predicted CPP (scaled) (all other variables held constant). Note that y-axes are not equivalent between figures.

9.4 Discussion

How do f0-based creak and acoustic measures of voice quality pattern according to age, gender and area in Scottish accents? In a large-scale analysis of 95 Glasgow, Lothian and Insular Scots, I found that found that f0-based creak was conditioned by linguistic factors (Duration, Speech Rate, Glottalisation, Phrase Position, Vowel, Contains /l,w,j/), and that Insular speakers were particularly creaky. Furthermore, voice quality outside of creak varied by linguistic factors (Duration, Speech Rate, Glottalisation, Aspiration, Pre-aspiration, Phrase Position, Vowel, Contains /l,w,j/, Contains Nasal, Contains Rhotic). In this non-creaky voice, female speakers showed evidence of tenser voicing than male speakers, evidenced by lower H1*–H2* and H2*–H4*, while younger speakers display a more modal-like quality than older speakers, shown in lower H1*–H2*, H2*–H4* and H4*–2kHz* as well as higher CPP. There was no significant difference by area. Here I discuss these methods and results.

9.4.1 Creaky voice

9.4.1.1 Duration

As stretch length increased, the log-odds of creak decreased. This reveals a limitation: As creak is separated from non-creak, new stretches are created, and these tend to be short if creak is present, as creak is rarely present throughout an entire stretch. This means that it is difficult to distinguish between speakers who creak often, but briefly (e.g. for glottals), from those whose creak forms a characteristically creaky voice quality.

This effect prompts a wider question: When does creak cease to be heard as a linguistically conditioned phenomenon (i.e. the listener hears only the effect of a glottal stop), and start to be heard as something that is characteristic of a speaker's overall voice quality? Future research might consider using a different unit of analysis, such as 10 ms frames following White et al. (2022). Given the scale of analysis, 10 ms frames would have been unworkable in the present study: 3,759 seconds of data would have resulted in 375,900 frames, which would have been computationally intensive to extract and manage. A useful avenue for future psychoacoustic research could therefore be to consider the threshold at which creak becomes part of a speakers voice quality, to obtain a practical unit of measurement.

9.4.1.2 Speech rate

Faster speech rate significantly decreased the log-odds of creak. This points to potential indexical usage of creak. As Pratt (2023: 9) discusses, slow speech is often connected to

low energy, while fast speech is often connected to high energy; Pratt (2023) connects both slow speech rate and use of creak to the expression of 'chill' affect. Future research could consider the potential affective use of creak in Scots further.

9.4.1.3 Glottalisation, phrase-final position and vowel onsets

A glottalable context - being adjacent to a /t/ that could be realised as a glottal - also significantly favoured creak. This is no surprise considering the pervasiveness of glottal stops in Scottish accents, as they can favour creak or be realised entirely as a portion of creak, with no full glottal closure. Potential pre-glottalisation contexts for did not appear to favour creak, which may suggest that pre-glottalisation and glottal stops as a form of /p/ and /k/ do not manifest as a period of low-f0 creak in Scottish accents.

Phrase-final position and vowel onsets also favoured creak. These effects, as with the effect of glottalisation, are in line with expectations based on previous research. However, no previous work appears to have considered the effect of linguistic factors on f0-based estimates of creak. Estimating these from the forced aligner rather than hand-coding for each linguistic effect shows the usefulness of f0-based estimation of creak as a coarse-grained method of estimating creak use in large amounts of data, even in spontaneous speech. This makes it possible to distinguish between linguistic uses of creak and potential other reasons, making it useful for sociophonetic research.

9.4.1.4 Creak in Insular Scots speakers

Creak was more prevalent in Insular Scots than in Glasgow Scots, while no significant difference was found between Glasgow and Lothian speakers. While Smith & Durham (2011) find that the use of local Shetland lexical, phonological and morphosyntactic features is lower among younger speakers, they note that the situation is highly complex with different speakers using local forms at different rates. One possible interpretation of the high use of creak in Orkney and Shetland is that creak may be a characteristically local voice quality that younger speakers may be maintaining, while their use of other certain local linguistic forms decreases. Future research could consider differences between Orkney and Shetland, which while grouped together here on the basis of shared historical and phonological characteristics, do form separate varieties; van Leyden & van Heuven (2006) find that native listeners can differentiate Shetland and Orcadian dialects only on the basis of prosody, for example. The functions of creak in Insular varieties is also a potential avenue for future research, as the present research only reveals that Insular Scots speakers are particularly creaky, leaving only speculation as to why.

9.4.1.5 Lack of age and gender difference

Based on the findings of the smaller corpus, I was not expecting to find gender differences in terms of creak. Consolidation of this finding on a wider scale has some important implications. Considering that Stuart-Smith (1999b) and Beck & Schaeffler (2015) found male speakers to be creakier using VPA, the lack of a gender difference may suggest a change over time. However, if this was the case, I would have expected to see a difference between older and younger speakers of the same gender. Alternatively, this lack of a difference could reflect the difference in methodology. As found in Chapter 7, f0-based identification of creak is more reliable in younger speakers and female speakers, problematising these findings. Future research should consider alternative automated methods of identifying creak, investigating how identification varies by speaker group to ensure that comparing speaker groups is viable.

9.4.1.6 Identification of antimodes and the use of REAPER

A major drawback in this portion of the research is the fact that f0 antimodes could not be identified for three speakers. Potential reasons behind this are discussed in more detail in Section 7.4, but it was hoped that using more data from each speaker would help to rectify this issue. Unfortunately, that was not the case.

It is possible that this method may have more success with a different f0 tracker. For example, the recently developed *h*f0 (Rengaswamy et al. 2021) combines deep-learning approaches with more traditional signal processing approaches to create an f0 tracker that purportedly performs well in the face of noisy signals. However, in the smaller corpus, hand-checked f0 measurement found examples of overlap between an individual's creaky and non-creaky f0 range, meaning that a purely f0-based method would never be able to accurately identify 100% of creak.

Though useful in the present study, f0-based coding of creak appears to have more potential when paired with other methods, such as in work by White et al. (2022). Doing so would potentially allow coding of creak that is better characterised by features such as f0 irregularity, damped pulses and multiple pulsing than low f0.

9.4.2 Non-creaky voice

Separating low-f0 creak from all other voiced speech allowed for analysis of the rest of speakers' voice quality. Using VoiceSauce, I measured H1*–H2*, H2*–H4*, H4*–2kHz* and CPP to estimate variation in phonatory setting in non-creaky voice according to linguistic and social factors.

		Phonation type				
		Whispery	Breathy	Modal	Tense	Tense whispery
Measure	H1*–H2*	→	↑	→	↓	→
	H2*–H4*	→	↑	↑	→	→
	H4*–2kHz*	→	↑	→	→	↑
	CPP	↓	↓	↑	↑	↓

Table 9.18: Patterns of different measures for different phonation types. A right arrow indicates that a measure tends to occupy moderate values for that phonation type, while an upwards arrow means it tends to occupy higher values, and a downward arrow means it tends to occupy lower values.

To recap, H1*–H2* can be broadly interpreted as a measure of glottal constriction, where higher values indicate a laxer phonation, and smaller values indicate tenser phonation. In the smaller corpus in this study, stretches with lower H1*–H2* were more likely to be tense rather than whispery stretches, and stretches with higher H1*–H2* were more likely to be breathy rather than whispery. Further, increasing H1*–H2* increased the predicted probability of a breathy stretch being coded as a higher degree of breathy voice.

H2*–H4* can also be interpreted as a measure of glottal constriction, but in the smaller corpus higher values for H2*–H4* corresponded to a modal rather than whispery quality, while even higher values corresponded to a breathy rather than whispery voice quality. Within breathy voice, increased H2*–H4* corresponded to an increasingly breathy-sounding quality. Lower H2*–H4* can therefore be taken as corresponding to a tense, whispery, or tense whispery voice quality, while intermediate values can be taken as modal, and higher H2*–H4* can be interpreted as breathy voice.

H4*–2kHz* is a measure of spectral tilt in a higher-frequency region of the spectrum. Zhang (2016a) finds that presence of a posterior gap in the vocal folds reduces excitation of harmonics in this higher-frequency region. In the smaller corpus, higher H4*–2kHz increased the probability of a stretch being breathy rather than whispery *and* the probability of a stretch being tense whispery voice rather than whispery.

CPP compares the amount of aperiodic noise and harmonic energy. Higher CPP corresponds to a nearer-to-modal voice, and lower CPP corresponds to a higher degree of noise, as in a more breathy or whispery voice. In the smaller corpus, higher CPP related to a tense or modal quality.

A summary of how these measures patterned together to create the perception of different phonation types in the smaller corpus is presented in Table 9.18. A right arrow indicates that a measure tends to occupy moderate values for that phonation type, while an upwards arrow means it tends to occupy higher values, and a downward arrow means it tends to occupy lower values.

9.4.2.1 Duration

As Duration increased, H1*–H2* and H4*–2kHz* decreased, while CPP increased. This suggests as stretches increase in length, phonation becomes more tense, with more glottal constriction and without a posterior glottal gap, and becoming closer to modal voice with lowering levels of noise. This may suggest that the longer the vocal folds vibrate, the more time there is for them to be positioned in a way to produce efficient voicing.

This may also be related to the analogous finding for creak. As the stretches are separated out into creaky voice and non-creaky voice, this makes creaky stretches shorter, simultaneously leaving non-creaky stretches as their full length, making these stretches more likely to be modal-like.

9.4.2.2 Glottalisation

In Glottalable contexts, which were immediately adjacent to a /t/ that could have been produced as a glottal stop, H1*–H2* and H4*–H2kHz* were significantly lower, while H2*–H4* was significantly higher. This suggests that these stretches were closer in quality to tense or modal voice. This finding is likely connected to the fact that creaky stretches have already been removed, leaving remaining stretches in Glottalable context likely to be tenser than adjacent stretches.

Preglottalisation contexts led to lower H4*–H2kHz. This suggests preglottalisation manifests in terms of closure of the posterior glottal gap in Scottish accents, rather than in a short period of creaky voice.

9.4.2.3 Aspiration

A preceding /p,t,k/ led to an decrease in CPP. This is consistent with predictions, and suggests that there is likely to be a small amount of aspiration noise from preceding voiceless plosives included within the voiced stretches identified by the automated procedure.

A preceding /h/ led to an increase in H1*–H2*. This is consistent with expectations. No increase in CPP accompanies this, potentially suggesting that /h/ in English is more consistent with a breathy rather than whispery voiced configuration, with little engagement of the laryngeal constrictor.

9.4.2.4 Pre-aspiration

In pre-aspiration contexts, there was significantly higher H1*–H2* and H2*–H4*, and significantly lower CPP. This strongly suggests that pre-aspiration occurs in Scottish accents, and that this manifests as a more open glottal configuration with more noise than in other portions of speech.

9.4.2.5 Phrase-final position

Stretches in phrase-final position showed significantly higher H2*–H4*. This can be interpreted in terms of the impact of separating out creak before the analysis: Creak has been taken away from phrase-final stretches, leaving more breathy or modal phrase-final stretches behind. This may also suggest that the form of creak present in phrase-final position is the lax creak that Slifka (2006) and Gobl & Chasaide (2000) have previously identified towards the ends of utterances. Research measuring spectral tilt in the phrase-final creak itself would be necessary to confirm this, however.

9.4.2.6 Vowels

Phrase-initial vowel onsets had higher H1*–H2*, H2*–H4*, H4*–H2kHz* and lower CPP, suggesting a more whispery or breathy onset. Again, this can be interpreted in terms of the fact that creak has been removed.

The other levels for this variable were containing both initial and non-initial vowels, and containing no vowels. Variation according to these factors suggests that segmental variation plays a key role in phonation.

9.4.2.7 Nasals, rhotics, /l,w,j/

Containing other consonants had various effects, but the most notable one is that containing any sonorant consonant decreased CPP. This should be taken into account in future acoustic voice quality research on spontaneous speech, as researchers may prefer to work only with vowels to allow for clearer interpretation of results.

9.4.2.8 Gender

Based on the findings in the smaller corpus, I had predicted that female speakers may have a more whispery voice quality, more tense voice quality and more tense-whispery voice quality, while male speakers might have a more modal and breathy quality.

Some evidence for this was found in the statistical analysis, with female speakers exhibiting lower H1*–H2*, suggesting that female speakers use a tenser voice quality than male speakers. Female speakers had a lower H2*–H4*, suggesting that female speakers may use less modal voice quality than male speakers.

Beck (1988), Stuart-Smith (1999b), Beck & Schaeffler (2015) that considered gender variation found a predominance of tense, whispery voice quality, and while Beck (1988) and Beck & Schaeffler (2015) found no significant differences in terms of whispery voice, Stuart-Smith (1999b) found that female speakers were more whispery. In this research, the smaller corpus study found again that Scottish voice quality is characterised by a tense whispery quality; The combination of tenser and less modal quality found here could potentially point to increased tense whisperiness in female speakers.

However, the gender differences identified here point more clearly to a tenser voice quality than to increased whisperiness. It is worth considering potential reasons for divergence from previous findings by Stuart-Smith (1999b). Stuart-Smith (1999b) also investigated stylistic variation between read speech and conversational speech, as well as differences according to social class, and found that these interacted. Stuart-Smith (1999b) found increased whispery voice was more apparent in female speakers in read speech, but in the present study, only conversational speech was considered and social class was not included as a variable; this could potentially explain the fact that female speakers were not clearly found to use a more whispery quality than male speakers. Furthermore, Stuart-Smith (1999b) found whispery voice was a marker of working-class speech compared to middle-class speech; in the present study, all participants are from a similar social class background, so potentially there is a high level of whisperiness generally which leaves little room for female speakers to be more whispery.

These findings also give counter-evidence to the claim that increased breathiness in female speakers is universal. In a study of gender variation in voice quality comparing Czech and Danish that considers this claim, Hejná et al. (2021) bring together their findings with those from a number of previous studies using H1*–H2* and H2*–H4* and note that the magnitude of gender differences varies between languages, suggesting that variation in breathiness is language-specific. Finding that H1*–H2* is lower in female speakers in Scottish accents allows us to take this claim further.

Firstly, it suggests that gender variation in glottal tension is also accent-specific. In British English more generally, Gittelson, Leemann & Tomaschek (2021) found that HNR and H1*–H2* were lower for male speakers than female speakers. Decreased H1*–H2* and non-significant differences in CPP therefore reinforce the distinct nature of Scottish voice quality from other British English varieties. Furthermore, it suggests that if there is a physiological influence that favours increased breathiness in female speakers, as suggested by Titze (1989), it is not deterministic.

In terms of H2*–H4*, the findings of the present study trend in the same direction as Hejná et al. (2021) for Czech and Danish, as well as for Chen et al. (2010) for English. As I found increased H2*–H4* to be most connected instances of modal rather than whispery voice in the smaller corpus, a potential physiological influence may still be at work, with greater H2*–H4* reflecting more modal-like phonation in male speakers, with full adduction of the vocal folds. We can then speculate that when a tense, whispery quality is used by female speakers, this requires increased laryngeal constriction relative to the production of a similar quality for male speakers, leading to lower H2*–H4*. However, H2*–H4* has still not been used as widely as H1*–H2*, and is not as well understood: Future research should investigate how it connects to underlying laryngeal configurations and auditory perceptions of voice in more detail, as well as whether this pattern exists cross-linguistically and across different accents.

The full picture of voice quality requires taking the relationship between multiple variables into account. Future research should consider using alternative modelling techniques in order to take these complex relationships into account, as modelling them separately here did not allow consideration of how the relationship between two acoustic measures varied according to independent variables. A tentative exploratory analysis, shown in Figure 9.15, suggests the relationship between H1*–H2* and CPP may vary by gender. Among male speakers, there appears to be a weak negative relationship, with CPP decreasing as H1*–H2* increases. This pattern suggests that noisier stretches are produced with an open glottal configuration, such that they would be consistent with breathy voice, while less noisy stretches are produced with a more closed glottal configuration, consistent with modal to tense voice. Meanwhile, female speakers show a weak positive relationship between CPP and H1*–H2*, with CPP increasing as H1*–H2* increases. Noisy tokens for female speakers appear to show a more closed glottal configuration, which would be more consistent with whispery or tense whispery voice, while less noisy tokens show a more open glottal configuration, consistent with a comparatively lax to modal voice.

Future research should consider the relationships between different acoustic variables more closely and investigate whether different types of 'noisy' voice quality can take on different social meanings.

9.4.2.9 Age

I expected to find lower CPP in older speakers, but did not find many age differences in the smaller corpus, so it is notable that age was the most relevant social factor in the statistical models. Younger speakers showed significantly lower H1*–H2*, H2*–H4* and H4*–2kHz* than older speakers, as well as significantly higher CPP. The divergence between the results of the PPA in the smaller corpus and the acoustic study

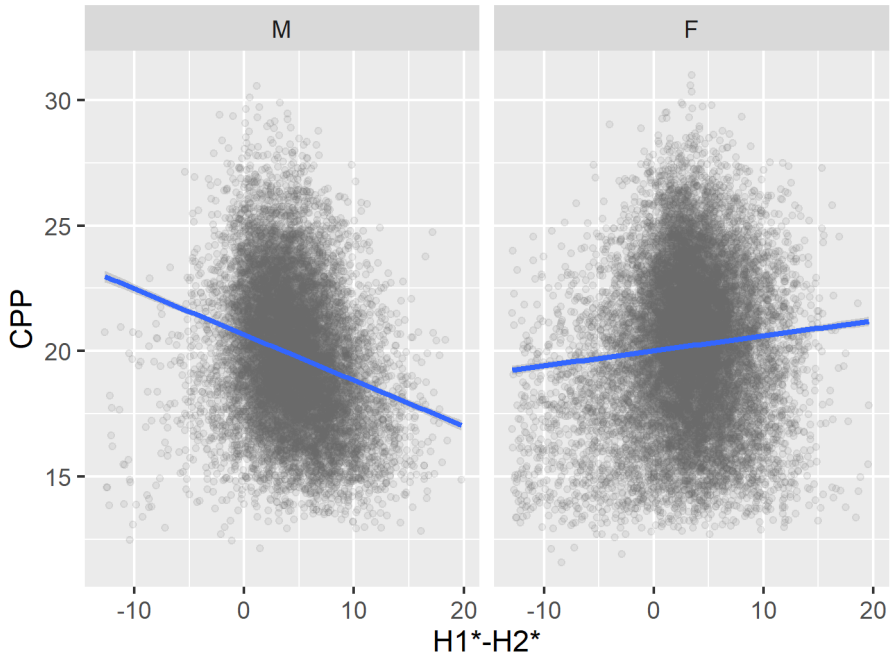


Figure 9.15: Exploratory scatterplot showing how the relationship between H1*–H2* and CPP varies by gender with smoothed conditional mean shown as blue line.

of the larger corpus points to these results being potentially related to the effects of ageing on the voice, rather than change in the use of phonation types. Taken together, these findings suggest that the effects of ageing on the voice are common across different Scottish accents, and supersede the differences between different phonation types: Older speakers produce a noisier, laxer quality across all of their voice quality, in a way that is difficult to capture in an auditory-perceptual scheme. This could be related to the difficulty of rating a continuous phenomenon on scalar degrees.

Previous research has mixed findings about the effect of ageing on the voice. In a review and meta-analysis of the effects of ageing on the voice, Rojas, Kefalianos & Vogel (2020) find that older individuals were rated higher for dysphonia, roughness, breathiness, strain and instability on clinical auditory-perceptual scales, and that this manifested in increased noise acoustically. However, increased breathiness and noise in older speakers is also not a clear-cut finding, with Rojas, Kefalianos & Vogel (2020) finding that increase noise among male speakers was more well-established than it was for female speakers. Similarly, Gittelson, Leemann & Tomaschek (2021) finds that for male speakers, HNR35 and H1*–H2* lowers with age, while for female speakers these increase, but finds no effect of age for CPP. Furthermore, Lee et al. (2016) warn against taking increased breathiness with age as a universal truth: They find tenser voice quality in elderly Korean women when compared to younger Korean women, as shown by decreased H1–H2 and H1-A1 with age. The findings of the present study, however, seem to show agreement with the general assumption of increased breathiness with age, as captured here in lower spectral tilt and higher CPP among younger speakers.

9.4.2.10 Lack of difference for area

Glasgow, Lothian and Insular Scots speakers exhibited different phonation profiles in the the smaller corpus, so the lack of differences between these groups in the larger acoustic study is surprising. This may be explained by the fact that whispery, tense and tense-whispery voice still predominated across accents, while distinctive voice qualities (breathy, modal and harsh voice) were rarer, potentially enough that differences are not notable acoustically.

Furthermore, while each phonation type showed a distinct acoustic patterning in the smaller corpus, these exist as complex relationships between acoustic measures, rather than absolute differences visible in one measure. Differences between accents may then not emerge when considering measure individually, and future research should consider how we can capture these relationships.

This finding also allows us to reflect on some of the advantages and limitations of auditory-perceptual and acoustic methods. The notable area differences found in PPA suggest that auditory-perceptual approaches may be better suited to identifying categorical differences in phonation types used more frequently in different accents, which are harder to identify using acoustic measures. Auditory-perceptual methods may also be more effective at identifying shifts away from a speaker or accent-specific baseline, such as cases where a speaker uses harsh voice, which would be difficult to identify using acoustic methods alone. Meanwhile, the continuous nature of acoustic methods means that they are well-suited to identifying linguistic influences on phonation, and to considering gender and age variance that persists across accents and phonation types.

9.4.3 Conclusion

Considering all of Part II together, we see that laryngeal voice quality in Scottish accents is characterised by whispery and tense whispery voice, but that it varies by social and linguistic factors. Insular Scots speakers use more creak than other groups, and voice quality varies by age and gender, with female speakers showing a tenser quality and younger speakers showing a tenser, more modal quality.

The influence of linguistic factors and age suggest that acoustic analysis may be well-suited to finding differences that are common across voice quality more generally, while auditory-perceptual analysis is well-suited to identifying groups of speakers who use particular phonation types more often. Future research should concentrate on how the complex relationships between different acoustic measures vary according to social factors within particular phonation types, further bridging these methods.

The analysis presented here shows the value of separating creak before proceeding

with subsequent acoustic analysis. If creaky voice had been considered alongside non-creaky voice in the acoustic analysis, I might have connected lower H1*–H2* among younger speakers and female speakers to creaky voice. This would have revealed little about what is happening in the voices of female speakers when they are not creaking (i.e. most of the time). Instead, we find that voice quality in Scottish female speakers is tenser, but *not* more creaky. This in turn problematizes iconicised links between female speakers and breathy voice. These findings also emphasise the role of a degree of auditory-perceptual analysis for interpretation of acoustic findings, as the PPA conducted on the smaller corpus allowed these findings to be interpreted with reference to descriptive labels for voices, and situated these within what we know about the auditory quality of Scottish voice quality.

One limitation of the corpus study is a lack of focus on the social meaning of this variation to speakers and listeners. Perceptual research may be useful to uncover how listeners perceive creaky voice in Scottish accents, while closer study of what is happening in a conversation when a speaker shifts may help to study how different social meanings may emerge. This corpus study provides a backdrop for research on Scottish voice quality. I turn now to one way that we might learn more about the social meaning of Scottish voice quality. In the next part of this thesis, I present a case study of the voice quality of Carrie, a Scottish transgender woman who uses her voice professionally in music and on the radio, and consider how she understands her own use of her voice. Through this in-depth analysis of a single speaker's voice and what it means to her, I hope to provide insight into the social meaning in Scottish voice quality and how this is constructed at the level of the individual.

Part III

Higher order indexicality: A linked phonetic and interpretative phenomenological analysis of a transgender woman's experience with her voice

Chapter 10

Introduction

In this thesis, I am interested in how macro- and micro- level analyses of voice quality relate to each other. In Part II, I considered this in terms of how auditory-perceptual analysis of voice quality, conducted using Phonation Profile Analysis, related to acoustic analysis of voice quality. Doing so allowed acoustic investigation variation in voice quality in Scottish accents by age, gender, area, and linguistic factors, which could be interpreted in terms of descriptive labels for voice quality. This revealed macro-level differences between groups, such as the use of creak by Insular Scots speakers, tenser phonation in female speakers than male speakers, and a tenser and more near-modal quality in younger speakers than older speakers. However, with 95 speakers, the investigation was unable to consider how the social meaning of certain phonation types emerged, and the acoustic focus meant that rarer but potentially socially meaningful phonation types, like harsh voice, were not considered as scale increased.

I therefore shift my focus here to considering how meaning emerges at the n+1st level (Silverstein 2003). The concept of indexical order, described in more detail in Section 3.2, provides a framework for understanding how a particular variable can take on new meanings by occurring in particular contexts and alongside other linguistic forms, allowing that meaning to be filtered through the meanings associated with those other variables and contexts and wider ideologies about them. Voice quality is particularly susceptible to this as it cannot occur in isolation, allowing it to become a 'semiotic hitchhiker' (Mendoza-Denton 2011). In Silverstein's (2003) framework, as n+1st order meaning emerges, speakers themselves become increasingly aware of it and able to shift their use of a variable to create a particular effect.

As Zimman (2017a) explores, the meanings that we attach to voices become more apparent for transgender speakers: As transgender speakers come out, their voices are re-contextualised as they begin to be gendered differently by others (Zimman 2017a). In these circumstances, covert ideologies about what male and female voices 'should' sound like become apparent, leaving trans individuals well-situated to give insight into

these ideologies from their experiences encountering them in Gender Identity Clinics, voice therapy sessions, and in other's reactions to their voices

Here, I present a case study of the voice of Carrie, a 50-year-old white, middle-class transgender woman from the West of Scotland. As a radio host, singer and writer who has voiced her own audiobook, she makes extensive use of her voice as a performance in her professional life. In her personal life, she uses her voice to talk to friends, navigate public spaces, and to move between 'two worlds', as she is out as a trans woman, but does not present as female around her children. I investigated the following question:

- How does Carrie understand her own experiences with her voice and use her voice to navigate different interactional contexts?

To investigate this question, I conducted and analysed a semi-structured interview with Carrie in accordance with the principles of Interpretative Phenomenological Analysis (IPA), an in-depth qualitative research method aimed at understanding how a person makes sense of a major life experience, in an attempt to understand how Carrie understands her own experiences of her voice as a trans woman. I also recorded her speaking to two different people, a close friend and an unknown interlocutor, to get a sense of how her voice changes between different situations. By combining acoustic analysis with IPA, I hope to provide you with insight not only into Carrie's voice, but into what her voice means to her and how voice quality can take on social meaning more generally.

I turn now to considering previous research relevant to this case study. I begin with research on trans speakers' voice quality to illustrate how working with transgender speakers can give insight into the way that voice quality takes on social meaning. I then consider the treatment of transgender speakers in sociolinguistic research more generally to justify the use of IPA in the present study. I then outline the key characteristics of IPA in more detail, and show what can be gained from an IPA study by exploring non-linguistic IPA research on trans people's experiences in the UK. I then give additional background on other aspects of trans people's experiences in Scotland, covering transition healthcare access and the Gender Recognition Act reform and backlash to it.

10.1 Trans speakers and the re-contextualisation of social meaning

Becker, Khan & Zimman (2022) define the term 'transgender' following the meaning that emerged from participants in their research: 'Transgender people are those whose

gender identities differ from the gender assigned to them at birth'. Transgender speakers often engage in processes of transition, which may involve both social aspects, such as changes to gender presentation and name, and medical aspects, such as use of hormones and accessing surgery. As Zimman (2017a: 361) explores, as trans speakers go through a transition process, the features of their speech and voice undergo a dramatic re-contextualisation process. This makes them particularly well-placed to give insight on the role of wider linguistic and non-linguistic stylistic context in creating social meaning. If, as Silverstein (2003) argues, social meaning emerges through a variable occurring in context of other variables and a wider social context, and in the ideologies that are attributed to that wider context, the same variant may be attributed different meanings when a speaker's gender is attributed to them differently by others. Zimman (2017a: 361) gives the example of this with regards to non-linguistic practice: 'a man who fawns over small children, takes up a lot of physical space on public transportation, compliments a woman's shoes, or playfully shoves a friend will be seen quite differently from a woman who does the same'.

Among transmasculine speakers, this re-contextualisation process often occurs over the course of taking testosterone. Drawing on data from ethnographic interviews, Crowley (2021) considers how two transmasculine individuals understand the gendering of their voices by others as they go through testosterone therapy. These speakers monitor how their pitch changes, and in doing so become aware of how other people's perceptions of them change as their voices lower in pitch. They find that interview participants report being given more space in conversations as their pitch drops and feeling that others now seem more likely to listen to them, and find that these interview participants draw on a sensory link between voice pitch and weight, deepness and heaviness to understand how they are attributed more social weight in conversations as they begin to be gendered as male by others.

Zimman (2012, 2017a,b, 2021) explores this process with regard to /s/, a variable which is often the focus of research on the social construction of gender because of how a higher-frequency /s/, often measured using Centre of Gravity (COG), can be associated with perceptions of vocal femininity, that are highly dependent of the context in which they occur (Pharao et al. 2014, Steele 2022). Zimman's (2012, 2017a) longitudinal ethnographic study of 15 transmasculine speakers in the San Francisco Bay Area in the early stages of testosterone therapy. Testosterone therapy is a form of hormone therapy which affects the physiology of the larynx and usually leads to a decrease in vocal pitch. He considers changes in f0 and the spectral Centre of Gravity (COG) of /s/, which relates to articulatory frontness and backness in /s/ and broadly connects with an auditory percept of how 'high' or 'low' the frequency of an /s/ sounds, where higher /s/ is often connected to perceptions of femininity. He found that f0 decreased with the amount of time speakers had been on testosterone, but that while some speakers showed no difference in /s/ COG, some speakers' COG decreased and one speaker,

Carl, saw an increase. He draws on observational qualitative data and ethnographic data to explain why this might be, noting that participants in his study discussed feeling more comfortable incorporating certain traditionally feminine characteristics into their gender presentation once they started to be perceived as men. He describes this happening in Carl's presentation during the course of the research. When they first met, Carl had an 'unremarkably masculine gender presentation that often took the form of slightly baggy jeans and t-shirts featuring pop-culture references' (Zimman 2017a: 359), but as his transition progressed, he began experimenting with a wider range of less normatively masculine clothing styles. Zimman interprets this as a form of stylistic bricolage (Eckert 2004, Hebdige 1979), arguing that a lower fundamental frequency and being gendered as male by others re-contextualises a high-frequency /s/, allowing it to be interpreted as indexing a queer or gay male identity (cf. Levon (2007), Pharao et al. (2014)), rather than cisgender, normative femininity. Working with transgender speakers, then, not only challenges the idea that these features will pattern together in normative ways, but allows consideration of how linguistic and social context that activates the social meanings of particular variables.

Zimman (2013) explores how this process works in terms of perception in a study on how transmasculine speakers' speech styles are often interpreted as gay-sounding, which focuses on the potential role of creak. He found both cisgender gay men and trans men use more creak than cisgender straight men, and that using more creak increased listeners' ratings of gayness, but that some of the speakers rated as the most gay-sounding used almost no creak. Considering this alongside other variables, he finds that cisgender gay and transgender speakers have disparate linguistic styles but are perceived as equally gay-sounding by listeners Zimman (2013) suggests that these results highlight the roles of stylistic context and multiple variables combining to create meaning, and that it is the inclusion of trans men in his study that offers the opportunity to consider that numerous phonetic styles can result in a similar percept and that there might be multiple avenues by which the attribution of a variable as being gay-sounding arises.

Some transgender people, particularly trans women and other transfeminine people, encounter the way their voice is re-contextualised as they take part in voice training. Voice training can be done independently using online resources (Kowalchuk 2020) or through attending speech and language therapy. As the most up-to-date version of the World Professional Association for Transgender Health (WPATH) Standards of Care (SOC8) (Coleman et al. 2022: s138) outlines, there are a wide variety of purposes for voice therapy, which includes enabling clients to communicate their sense of belonging in terms of gender identity in everyday encounters in a manner that matches their self-presentation. The SOC8 (Coleman et al. 2022: s139) recommends assessing an individual's desired communication function in a way that presents them with respect and autonomy in decision-making, rather than involving pre-determined goals.

In practice however, much of the research carried out in this area involves a number of ideologies about what male and female voices should sound like. Reviewing ideologies that occur within this body of research, Zimman (2012: 97-116) identifies patterns of voice therapy research naturalising gender differences between male and female voices, viewing gender differences in the voice as directly related to physiology, pathologising trans voices and framing trans women's voices as needing intervention by speech therapists, who are positioned as the arbiters of acceptable femininity. Similar issues are revealed in research on trans people's experiences with voice therapy. Thomson, Baker & Arnot (2018) looked at people's experiences with accessing speech and language therapy through a gender identity clinic in Scotland and found that only 43% felt that clinicians had explained the different options available to them, and only 57% felt that their personal needs and preferences were taken into account. They give examples of several people's experiences accessing speech and language therapy, with one participant saying that while the therapy was excellent, they had to stop attending because they 'felt under pressure to dress female', with another participant saying that 'The speech therapist told me that I was too feminine and that I spoke too much like a girl and that she couldn't help me' (Thomson, Baker & Arnot 2018: 85). In practice then, speech and language therapy is likely one place where trans people encounter ideologies about gender differences in the voice, potentially giving them a unique insight into this area. While I leave a great majority of the voice therapy research to one side, due to a lack of focus on social meaning of voice quality, I note that it reveals that the use of breathy voice is often a target in feminising voice therapy, alongside an increase in pitch and resonance (Davies, Papp & Antoni 2015).

Overall, work on trans speakers reveals how social meaning is highly dependent on context, and that trans speakers themselves are well-situated to reflect on this meaning as their voices as recontextualised as they go through a process of transition and begin to be gendered differently by others.

10.2 Interpretative Phenomenological Analysis

As well as analysing how Carries's voice quality differs between different contexts, here I adopt Interpretive Phenomenological Analysis (Smith 1996, Smith, Flowers & Larkin 2022), a method initially developed for use in health psychology (Smith 1996). IPA aims to understand how participants make sense of their major life experiences. Here I argue that IPA is well-suited to close investigation of participants' experiences with their voice and how they understand them, and the relationship between this and their voice quality.

10.2.1 What is qualitative research?

The distinction between qualitative and quantitative research is sometimes seen as clear-cut: Quantitative research supposedly deals with numbers and close-ended questions and responses, while qualitative research deals with words and open-ended questions and responses (Creswell 2017: 3). However, as Creswell (2015) argues, these are perhaps better framed as being different ends of a continuum, with some research sitting closer to the quantitative end, and other research sitting closer to the qualitative end. Mixed-methods research, incorporating some qualitative and some quantitative elements, sits in the middle. Different aspects of research — the aims, data, methods, analysis, interpretation, approach to research and resulting outputs — can also be more characteristically qualitative or quantitative in nature (Ormston et al. 2013: 4)

Part II mostly relied on quantitative methods to investigate variation in voice quality, which was effective for its macro-scale focus. However, research question of this part of the thesis, which is focused on Carrie's understanding of the social meanings attributed to her voice, might be best understood as qualitative in nature. According to Ormston et al. (2013: 3), qualitative research includes a wide range of approaches and methods across different disciplines, but certain characteristics give qualitative research its distinctive character, and one of these characteristics is that the aims and purposes of qualitative research are typically, 'directed at providing an in-depth and interpreted understanding of the social world of research participants by learning about the sense they make of their social and material circumstances, their experiences, perspectives and histories' (Ormston et al. 2013: 4).

10.2.2 Principles of IPA

Smith, Flowers & Larkin (2009: 1) describe IPA as an approach to qualitative research that is committed to the examination of how people make sense of their major life experiences. It provides a theoretical underpinning, guidance for designing semi-structured interviews, and a common set of principles for analysing data and interpreting results which were drawn up with this commitment in mind. It was first presented in Smith (1996)'s paper, which argued for IPA as a qualitative method in psychology that could capture people's experiences while also being able engage in dialogue with existing psychological research, but is now used beyond psychology. While few studies in linguistics have used this specific approach, Farr et al. (2018)'s study of migrant mothers' experiences of speaking their heritage language with their children and how this relates to their ethnic identity shows that there is a place for IPA in linguistic studies concerned with how participants understand the relationship between their own language use and identities.

As well as potentially being useful in linguistic studies more generally, IPA is also well-suited specifically to the study of style-shifting and vocal change in LGBTQ+ speakers. As discussed in last section, studies looking at style-shifting and vocal change in transgender speakers specifically (Zimman 2017a, Jones 2022, Zimman 2021) have collected some kind of qualitative data, either through ethnography, sociolinguistic interviews, or in the form of conversational data where participants happen to talk about their identities or gender experiences. This qualitative data then allows the researcher to explain parts of their findings in terms of themes that arose in the interviews or conversations, and these studies, particularly Zimman (2017a), have shown that a qualitative angle can be especially useful for examining how a speaker's linguistic behaviour relates to their broader life experiences. However, because this data is generally collected with a sociolinguistic or phonetic analysis of the data in mind, rather than a more explicitly qualitative focus, collection of qualitative data is conflated with collection of speech data, and analysis of qualitative data lacks a clear framework. This means that the way the qualitative data is analysed is not always transparent and systematic. IPA, on the other hand, guides the way that the research is designed as well as the way that the qualitative data is analysed, and allowing the researcher to focus on how participants' understanding of their own experiences relate to their linguistic practices from the outset of the research.

IPA is therefore suited to this research for several reasons. One reason is its idiographic focus, meaning that it is concerned with the particular (Smith, Flowers & Larkin 2009: 29). IPA's commitment to the particular exists on two levels: Firstly, it committed to a detailed, in-depth analysis, and doing this in a thorough and systematic way. The advantage of this for the present study is that IPA gives a clear, systematic framework for the analysis of qualitative data. Secondly, IPA is idiographic in the sense that it is concerned with how particular experiential phenomena (in this case, the experiences with the voice), have been understood from the perspective of particular people in a particular context (here, the perspective of Carrie, a trans woman who came out later in life and uses her voice extensively in her work). In particular, Smith, Flowers & Larkin (2009: 33) note that IPA is interested in experiences that are of particular significance to the people experiencing them, such as major life transitions. Style-shifting is not a particularly significant experience in itself; however, the way that IPA understands significant life experiences is through particular, everyday moments that become significant as part of a major life experience (Smith, Flowers & Larkin 2009: 2). In the present study, the particular instance of style-shifting studied forms one of these experiences that can then help to make sense of how a participant understands a major life transition such as coming out, or going through social or medical transition.

10.2.3 Comparison to other qualitative approaches in sociolinguistics and phonetics

IPA is a relatively new approach in sociolinguistics, particularly in studies that also consider phonetic data. However, many sociophonetic studies involve qualitative aims, data, methods, and interpretation, and often use sociolinguistic interviews and/or ethngraphy to approach these.

Certain principles of IPA are shared with linguistic ethnography, an approach with a long history in sociolinguistics that developed out of work by Hymes (1968, 1972) and Gumperz (1982), which looks at how everyday actions are embedded in wider social contexts and structures (Creese & Copland 2015: 12). In common with linguistic ethnography, IPA takes an interpretive approach, attempting to look at these actions from the participants' point of view and engaging in analysis with an understanding that analysis involves the researcher's interpretation. However, in ethnography, the researcher also engages in first-hand interpretation of the action, through observation of participants; This is where IPA begins to diverge, as in IPA the researcher only engages in interpretation of the participant's account of the action. Furthermore, in ethnography, the researcher's interpretation of the event serves to 'make the familiar strange' (Erikson 1986: 121), allowing analysis of everyday interactions and routines that participants themselves may not pay attention to because of their familiarity. In contrast to this, IPA is interested specifically in what happens when these everyday experiences take on a particular significance to the participants themselves (Smith, Flowers & Larkin 2009: 1), and puts the participant in the position to reflect on these everyday experiences themselves. IPA's approach is more suited to the present study, because the participants' own awareness of their own voices and experiences of vocal change and shifts and how they interpret these themselves is central to the research question.

Sociolinguistic interviewing aims to incorporate aspects of ethnography in a one-to-one interview that serves as data about the participant's life experiences (Labov 1984). There is a degree of flexibility in this too, with different researchers adapting sociolinguistic interviewing techniques to obtain information about particular topics of interest, as well as allowing the interview to be guided by topics that are of interest to the interviewee. In the present study, I favour the use of IPA over sociolinguistic interviewing as I find it advantageous to separate the collection of *speech data*, which will ultimately be analysed quantitatively, with the collection of *qualitative interview data*, which will be ultimately analysed using a qualitative approach. Collecting these two types of data simultaneously limits the flexibility of both the qualitative questions that can be asked, and the type of speech style that can be elicited: Questions that are optimal for eliciting a particular speech style may not be optimal for investigating

the participant's perspective on a particular issue. Furthermore, there is no unified framework or set of principles for conducting a qualitative analysis of sociolinguistic interviews, leaving sociolinguistic interviewing susceptible to cherry-picking of qualitative data by researchers.

Only one previous study has been identified that combines phonetic analysis with IPA. Saleem (2020) considers the experience of 10 Punjabi trans women in Lahore, Pakistan, combining phonetic analysis of fundamental frequency and vowel formants with IPA of semi-structured interviews. The IPA revealed that participants used a variety of strategies to adapt their voices, ranging from taking herbal supplements, to copying actors and singers, to watching YouTube videos, to formal speech therapy. Participants had diverse feelings and experiences surrounding their voices, ranging from acceptance, seeing the voice as adaptable and feeling as though their voices were congruent with their presentation, to issues with self-confidence because of others gendering them as male because of their voices.

The study reveals a potential difficulty of integrating IPA and phonetics, however, as the nuance in the way that participants themselves talk about their own voices is not reflected in the how the phonetic analysis is reported. For example, two participants mentioned the way that nuances in their identities related to them accepting their voices as they were, with one saying, 'When I started identifying more as trans woman, rather than a female, I became more comfortable with wherever my voice is at', while another said, 'honestly I do not want to sound exactly like a female'. Others talked about putting in significant effort to change their voices, which some felt had resulted in a change in how confident they felt and/or how they were gendered by others, but for others, they felt there were still aspects of their voice they wanted to work on in order to feel confident or change the perceptions of others.

In contrast to this account of the diverse range of perspectives that trans women who took part in the study had about their voices, the phonetic analysis proceeds by comparing trans women's voices to those of cisgender men and women, and evaluating who they sound most similar to. This appears disconnected from the more nuanced IPA account of each participant's experience. This study shows some of the potential of IPA, but also the challenge of attempting to pair it with classic sociophonetic analysis.

Despite these potential challenges, IPA still offers promise of providing a framework for conducting qualitative research in the present study. In particular, IPA research on transgender people's experiences in the UK shows its potential for showing the nuances in trans people's experiences. Drawing heavily from IPA research, I turn now to the backdrop to the present case study, outlining some of the current issues facing trans people in the UK and Scotland specifically.

10.3 Transgender people's experiences in the UK

The backdrop to the present research is a volatile moment for trans people in the UK and particularly in Scotland. As Pearce, Erikainen & Vincent (2020) explore, trans issues are at the centre of a public discourse circulating online, in mainstream media, and in academic publications. As Pearce, Erikainen & Vincent (2020), Zanghellini (2020), Hines (2020) outline, in 2017 the UK Government announced plans to hold a public consultation on reforming the 2004 Gender Recognition Act (GRA), which led to a subsequent backlash and debates over relations between trans rights and feminism. The proposed reforms affected the process that allows trans people to obtain a Gender Recognition Certificate (GRC), a legal document which allows someone to change the gender on their birth certificate. Under the 2004 Act, obtaining a GRC is a lengthy process requiring the applicant to provide evidence of living in their acquired gender for at least two years, diagnosis of gender dysphoria, a report of any medical treatment they have received, and if married, agreement from a spouse. Potential reforms involved removing some of these requirements and replacing them with a statutory declaration of an individual's self-identified gender.

Discussed in more detail by Pearce, Erikainen & Vincent (2020) and Zanghellini (2020), opposition to reforms centred around the idea that self-identification is dangerous and would pose a threat to women's rights, by allowing 'men' into women's spaces such as toilets, changing rooms, and crisis centres, where 'men' is used as a catch-all term to include trans women and non-binary people assigned male at birth and hypothetical cisgender men impersonating trans women. Zanghellini (2020) notes how these assumptions are firstly flawed on a practical level, as access to women's spaces is not granted on the basis of the sex on someone's birth certificate, while Pearce, Erikainen & Vincent (2020) situate the backlash in a longer history of positioning trans women as a threat to cisgender women's safety and drawing on the language of women's liberation to frame opposition to trans rights as a feminist (e.g. Raymond 1979).

Although the consultation provided evidence of widespread support for reforms, with a majority of respondents supporting removal of the requirements outlined above (King, Paechter & Ridgway 2020), the UK Government decided against implementing reforms (Parker 2020). Similar set of reforms were consulted in Scotland, eventually passing through the Scottish Parliament in December 2022. This led to wider discourses over potential reform becoming concentrated on Scotland, an issue compounded by debates over the Scottish Hate Crime Bill (Montiel-McCann 2022) and how to best record information about sex and gender in the 2022 Scottish Census (English 2022) and by the prominent opposition to reforms by author J.K. Rowling (e.g. Rowling 2022). Opposition led to the reforms ultimately being blocked by the UK Government in January 2023, shortly before data collection began for this part of this project.

Trans individuals in the UK also faced increasing difficulty accessing transition-related medical interventions, such as hormone replacement therapy and gender-affirming surgery. As R. Pearce (2018: 51-80) outlines, the process behind accessing trans healthcare in the UK involves getting a diagnosis of gender dysphoria following referral to a Gender Identity Clinic (GIC). Waiting times to be seen at GICs, however, are lengthy, with Harrison, Jacobs & Parke (2020) finding that individuals wait an average of 136 weeks to be seen at a GIC. However, COVID-19 led to many appointments being cancelled or postponed, meaning that waiting lists are now far longer.

R. Pearce (2018: 19-50) traces the origins of the creation of the 'gender dysphoria' diagnosis in process of pathologisation of trans identities, where deviation from gender norms is framed as illness. R. Pearce (2018) argues that while the existence of gender dysphoria as a diagnosis ensures that treatment is available, it also creates a distinction between trans people who require access to medical transition and those who do not, requiring trans people seeking to access to prove their themselves and creating a class of gender identity experts who act as gatekeepers of this care. R. Pearce (2018: 5) notes that there is a paucity of research on trans people experiences navigating this system in the UK, with studies on trans health instead being rare outside of medical journals. Trans people's experiences in this recent UK context have been explored in R. Pearce's (2018) own work, and in a number of recent IPA studies.

Harrison, Jacobs & Parke (2020) considers the experience of trans people in the UK seeking medical transition through the lens of gender dysphoria diagnosis. They conducted an IPA study of eight participants' experiences with transition and navigating the referral and treatment process. The first theme explores participants' dissatisfaction with the process of accessing treatment on the NHS, with waiting periods and lack of communication from GICs resulting in anxiety and frustration. In the second theme, they explore participants' process of searching for acceptance. Participants saw acceptance from families as important, but perceived that tension arising upon discussing their transition led to breakdowns in familial relationships. Furthermore, participants discussed challenges in the community, reporting experiencing verbal abuse and threats of violence, which for some participants made them fearful to leave the house and led to social withdrawal. Meanwhile, relationships with friends were less affected, and this support had a positive effect. Finally, they explore how the long process of accessing medical treatments negatively impacted participants' mental health, with participants also facing barriers to accessing mental health treatment during their transitions.

Mills et al. (2023) critique Harrison, Jacobs & Parke (2020) on the basis that their work centres around the concept of gender dysphoria, thereby pathologising trans people's experiences. They instead centre the role of the social environment in distress experienced by trans people. Mills et al. (2023) focus specifically on the experiences of transmasculine individuals within this system, seeking to understand how trans-

masculine adults in the UK understand identity development within the context of seeking transition healthcare. Their IPA research with 12 transmasculine individuals resulted in the disucssion found three superordinate themes: Conceptualising Stages of Transition; NHS Communication and support; and Medicalisation, Power, and Non-Disclosure. In the first theme, Mills et al. (2023) discuss how participants conceptualise transition in 'stages' of identity development such as self-acceptance, searching for answers, and coming out, but also emphasise that some stages are ongoing and a lifelong process. They also conceive of convincing gender identity services of their authenticity as a battleground, framing this as fight with multiple stages. In the second theme, they explore the impact of isolation and lack of communication during waiting periods on participants' well-being and mental health. Finally, they explore the power participants perceived the gender identity services to hold, and how fears of treatment being withheld led to participants hiding mental health struggles from gender identity services and how participants felt a pressure to conform to clinical expectations of binary gender norms to access treatment.

Eisenberg & Zervoulis (2020) focus specifically on the experiences of older trans women in the UK, allowing more attention to the particularities of this experience, such as struggles in relationships with partners, children and friends, and having grown up in a period where being trans was still highly stigmatised. They consider how six trans women over 35 who began their transitions in adulthood understand the development of their identities. Participants conceptualised their transitions in five broad stages: feeling different, conforming, exploration, coming out & transitioning and authenticity. Participants shared a sense of feeling different from an early age, but because of lack of exposure to information about trans issues, were unable to understand these feelings. Rigidity of gender roles during participants' childhoods led to participants conformed to expected male behaviours in an attempt to find acceptance from others, which some participants found contributed to episodes of depression and anxiety. They recount how participants report exploring feelings sporadically in early life, before being able to explore their feelings more fully upon information about trans people becoming more widely available. Coming out to friends and family then led to redefining their place in relationships and society. One participant talked about living publicly as a woman, but living privately as her gender assigned at birth around her children to maintain her identity as a 'father', while other participants found that coming out led to severing ties with family entirely. They report that coming out allowed participants to find a new sense of authenticity, allowing them to enter new relationships that they saw as more genuine, but that lack of support from existing relationships continued to affect participant well-being.

Together, the current context of debates over GRA reform in the UK and IPA research on trans people's experiences in the UK reveal key aspects of experience that may be relevant to the present study. Firstly, GICs and gender dysphoria diagnoses

affect whether participants are able to access interventions that affect their voice, notably, testosterone therapy and voice therapy, which are often the focus of previous work on trans voices. Less directly, this power that this system holds often trans people wishing to access treatment (Mills et al. 2023) may mean that participants feel pressured to conform to gendered expectations in terms of their voice in order to access treatment. Secondly, previous research on productions of /s/ among LGBT speakers in rural California (Podesva & Hofwegen 2015) has identified that the sociopolitical landscape can pressure speakers to adhere to gender norms in terms of their speech. The wider negative environment for trans people in the UK, which manifests in public discourse on trans people (Pearce, Erikainen & Vincent 2020), public harassment (Harrison, Jacobs & Parke 2020), and disruption of personal relationships (Eisenberg & Zervoulis 2020), could therefore be relevant to trans people's experiences with their voices.

Overall, IPA shows promise for gaining insight into how the participant in this case study understands her own experiences with her voice, and how this relates to her use of her voice more widely. I move now to outlining the methods used in the present case study, where I integrate IPA with qualitative and quantitative phonetic analysis to illustrate how Carrie's understanding of her voice and use of her voice across different contexts relates to the social meaning attributed to voice quality more generally.

Chapter 11

Methods

11.1 Methods

In this case study, I investigated how a participant understands her own experiences with her voice and uses her voice to navigate different interactional contexts using a mixed-methods approach that combines Interpretative Phenomenological Analysis with qualitative and quantitative phonetic analysis.

I initially aimed to work with three participants, but due to time constraints and the richness of IPA data, I present a case study of a single participant. I recruited this participant via a sampling survey that created a pool of potential participants.

The case study therefore involved three stages of data collection:

- The sampling survey
- A qualitative interview designed in accordance with the principles of Interpretative Phenomenological Analysis, investigating how Carrie understands her experience with her voice and how she uses it, referred to as the 'IPA interview'
- Carrie being recorded speaking to two different interlocutors, referred to as the 'recorded conversations'

The data was then analysed as follows:

- The IPA interview was analysed following the principles of Interpretative Phenomenological Analysis
- The recorded conversations were analysed in a linked auditory-perceptual and acoustic analysis that built on the methods developed in Part II
- An integrated qualitative-acoustic analysis considered instances where Carrie discussed and produced different 'voices' that she used in her life

The research received ethical approval from the University of Glasgow ethics committee, after being escalated by the College of Arts ethics committee. The ethical approval process involved conducting a Data Protection Impact Assessment, which was discussed with the University of Glasgow's Data Protection and Freedom of Information Office.

11.1.1 Sampling survey

The survey functioned as a sampling and recruitment tool for the recorded conversations and IPA interviews that followed. The initial plan to work with three participants

meant that finding a homogenous sample with similar demographic characteristics and similar experiences with their voices was important (Smith, Flowers & Larkin 2009: 48-49).

The survey asked questions about participants experiences with their voices during their transitions, and contained the following sections:

1. Experience with your voice in different situations
2. Experience with trying to change your voice on your own
3. Experience with transition-related healthcare
4. Demographic questions and in-person interest

Each section contains a combination of multiple choice and open-response questions aimed at providing wider context to the interviews. Initially, the plan was to analyse these responses in full using thematic analysis to provide a wider context to the study. However, due to time constraints, the survey was not analysed in full.

The survey was conducted via Qualtrics (Qualtrics 2005).

11.1.1.1 Survey participants

The survey was aimed at people over the age of 18 who considered themselves to be transgender and from and currently living in Scotland. It was open to people who may not specifically identify with the term 'transgender', but who do not identify wholly or exclusively as the gender they were assigned at birth. It was specified that this included, for example, people who identify as trans, transsexual, or as having a transgender history, as well as people who identify as non-binary, genderqueer, agender, or otherwise outside of the gender binary. Participants were recruited through social media posts shared by myself and the Scottish Trans Alliance. To compensate participants for sharing their experiences, participants were offered a £5 voucher for Tesco or Category Is Books, a local LGBT+ bookstore, or a £5 'thank you' donation to be be split between LGBT Health and Wellbeing and the Scottish Trans Alliance.

11.1.1.2 Survey content

11.1.1.3 Spam and bot responses

The £5 voucher incentive attracted a large number of spam and bot responses. The following procedure was developed for identifying authentic responses:

- Checking the QRecaptcha Score generated by Qualtrics, where scores closer to 0 are more likely to be bots, and scores closer to 1 were considered more likely to be real people). Responses with no QRecaptcha Score or a score less than 0.8 were determined to be likely bots
- Disregarding responses in other languages (Latin and Chinese)
- Checking for repetitive or nonsensical answers. For example, a large number of responses, deemed to be bots, gave their pronouns as '666', 'I was born as a man, but I prefer you to call me a woman' or 'transgender'.
- Checking for incompatible answer to multiple choice questions. A large number of responses, deemed to be bots or spam answers, indicated that they had received voice therapy, vocal fold surgery, and testosterone therapy, and remedial speech and language therapy related to their transition, or indicated the same answers when answering which forms of transition healthcare they had already accessed and hoped to access in the future
- Disregarding batches of highly similar responses received within 10-minute time spans

11.1.2 Introducing Carrie

Carrie is a 50 year old woman from Ayrshire who indicated her interest in taking part when completing the survey. In the survey, she describes her voice as 'a very low, resonant voice that's a bit scratchy because I talk/sing a lot.' However, she uses her voice differently in different situations: 'When I'm presenting fully feminine I tend to focus on my vocal pitch and raise it'. She mentioned that she would 'like a more stereotypically female voice - more vocal variety as well as higher pitch and less resonance'.

She reflected on some of the issues she faced in her attempt to use her voice in this way: 'One of the big problems for me has been code-switching: I didn't present female on the school run or around my kids so I was effectively living with two different voices, a male and a female one, and switching depending on where I was/who I was with. That makes it really hard to just do it automatically'. She also gave insight into some of the reasons that this was important to her: 'The voice is one of the most frustrating things because people instantly decide your gender when they call you or hear you - if someone hears me rather than sees me first, they'll instantly gender me male even if I'm using a more feminine voice. It makes getting car insurance something of a challenge because they don't believe I'm the policyholder (I have a female name)'.

Initially, the plan was to recruit three participants to take part in the in-person

portion. However, the impact of COVID meant that there were ethical concerns about collecting in-person data which remained until 2022, delaying plans for data collection. Furthermore, as the project involved working with transgender participants, conducting an in-depth interview involving sensitive topics, and sampling participants via a survey which raised data protection concerns, this lead to a lengthy ethical approval process involving a Data Protection Impact Assessment. These delays meant that only two participants were recruited, and only the data from one participant, Carrie, was able to be analysed here.

The two participants recruited were Carrie and Eilidh (both participants chose to be identified by their first names rather than a pseudonym). Carrie and Eilidh were recruited using the via the survey. They were selected on the basis that they were both trans women from the West of Scotland whose survey answers suggested that they had both been actively engaged a process of altering their voices (both mentioned a variety of activities they had engaged in in an attempt to change the sound of their voices and reflected on the process), that they had similar goals in terms of what they wanted their voices to sound like, and both talked about singing. Eilidh's data is not considered here, due to time constraints.

11.1.3 IPA interview

11.1.3.1 Interview procedure

Carrie took part in a semi-structured interview: While the interview schedule given in Appendix C.2 contains a range of topics that should be covered to enable the research question to be answered and list of sample questions and prompts to enable these topics to be discussed, the interview was guided by Carrie and aspects of experience that emerged as important to her.

The interview aims to cover the following areas:

- Describing the experience of voice change
- Voice and interpersonal relationships
- Voice and identity

Carrie was asked open questions in each of these areas, followed by prompts and probing questions to encourage her to share examples, reflections, and her own understanding. Carrie received the interview schedule in advance and was informed that the questions were flexible, allowing her to disclose any topics to avoid and make an informed decision on taking part.

11.1.3.2 Analysis following IPA

Central to IPA is an analytic focus on directing the researcher's attention towards participants' understanding of their own experiences. Its approach to data analysis can be characterised by a set of common processes, principles and strategies, explained in detail in Smith, Flowers & Larkin (2022). The interview with Carrie was analysed following these guidelines, and as a novice IPA researcher I stayed close to the steps that Smith, Flowers & Larkin (2022) set out. These steps are as follows:

1. Transcription, listening, reading and re-reading
2. Exploratory noting
3. Developing experiential statements
4. Searching for connections across experiential statements, to develop Personal Experiential Themes (PETs)

The steps that Smith, Flowers & Larkin (2022) set out also include 'Moving to the next case' and 'Looking for patterns across cases'. However, as this is only a case study, this was not necessary here.

11.1.3.3 Transcription

I transcribed by hand using Transcriber (Barras 2002) to create a time-aligned transcript. In aimed to produce a verbatim transcript, so that false starts and hesitations were recorded, so that I was able to draw on these in the analysis if needed. Marked pauses were also recorded, though unlike in other transcription systems like Conversation Analysis, I did not record the length of the pause.

11.1.3.4 Exploratory noting

I conducted the exploratory noting on a physical transcript of the interview with wide margins. This process involves close reading of the transcript and commenting on semantic content, language use, and how the participant talks about aspects of the experience that matter to them. I intially followed Smith, Flowers & Larkin's (2009) suggestions differentiating between types of comments by colour. Descriptive comments, in black, described the content of what Carrie was saying, while linguistic comments, in purple, commented on the language used, and conceptual comments, in pink, often took the form of questions about what the participant's words meant and began to offer my own interpretations. I also included phonetic comments, in green,

that reflected on instances of local-style shifting, such as a shift in phonation. As I progressed through initial noting I found that changing between pens took me out of the flow of analysis, and I therefore made most comments in black. During the process, I re-listened to the interview twice, and re-read the transcript itself without audio several times. Listening to the interview recording helped me to slow down in my analysis, to take the interview as a whole and think about how it progressed, and to connect Carrie's words to her voice. Noting on the transcript without listening helped me to begin to think about connections between different parts of the interview more easily, as I could look at different parts of the interview side-by-side and consider particular sections in more detail.

An example of some of the initial noting can be seen in Figure 11.1, which shows a scan of one of the pages of the interview transcript.

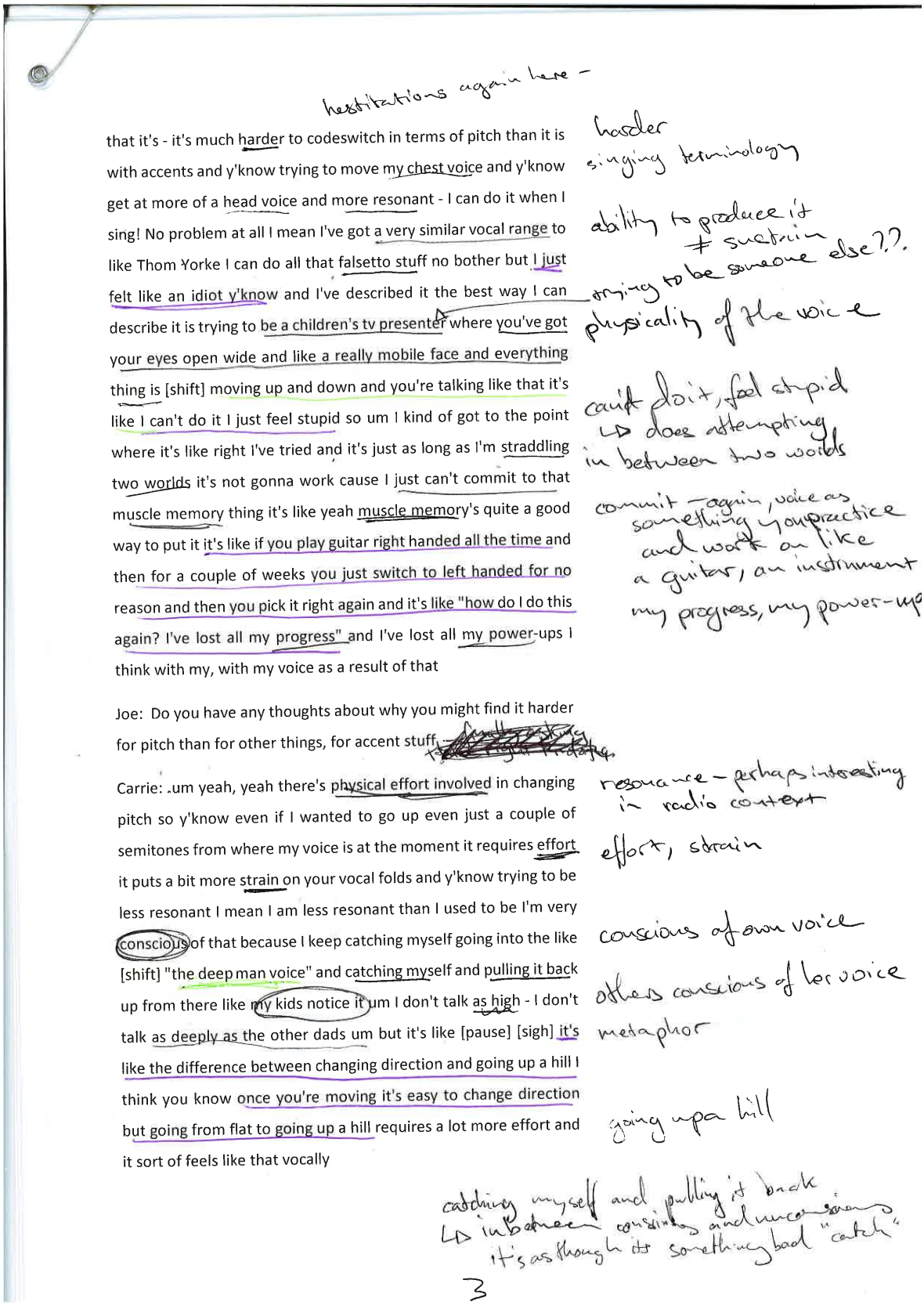
11.1.3.5 Developing Experiential Statements

I then moved to the other margin of the transcript, and began to draw on the exploratory comments to develop Experiential Statements. The aim of this part of the analysis is to maintain the complexity and important features of the exploratory notes while reducing the volume of detail, constructing a statement that summarises the important parts of the initial notes. Smith, Flowers & Larkin (2022) suggest moving away from working with the transcript and towards working with the notes in this step. As I began to construct Experiential Statements, I discovered my initial notes often lacked descriptive comments, and included more conceptual and linguistic comments. Because of this, I returned to the transcript during the process of developing Experiential Statement to write more exploratory comments, and these two stages were not entirely distinct in my analysis.

11.1.3.6 Searching for connections across experiential statements

This step involves organising experiential statements into a structure and explore relationships between them by typing up and printing out the statements, then cutting them up and laying them out on a table. This allows the statements to be organised into clusters. I ended up with eleven main clusters of statements organised on the table, with some statements placed in between clusters where they demonstrated a relationship between two clusters.

Figure 11.2 shows this process.



hesitations again here –

that it's - it's much harder to codeswitch in terms of pitch than it is with accents and y'know trying to move my chest voice and y'know get at more of a head voice and more resonant - I can do it when I sing! No problem at all I mean I've got a very similar vocal range to like Thom Yorke I can do all that falsetto stuff no bother but I just felt like an idiot y'know and I've described it the best way I can describe it is trying to be a children's tv presenter where you've got your eyes open wide and like a really mobile face and everything thing is [shift] moving up and down and you're talking like that it's like I can't do it I just feel stupid so um I kind of got to the point where it's like right I've tried and it's just as long as I'm straddling two worlds it's not gonna work cause I just can't commit to that muscle memory thing it's like yeah muscle memory's quite a good way to put it it's like if you play guitar right handed all the time and then for a couple of weeks you just switch to left handed for no reason and then you pick it right again and it's like "how do I do this again? I've lost all my progress" and I've lost all my power-ups I think with my, with my voice as a result of that

harder
singing terminology

ability to produce it ≠ suction

trying to be someone else??

physicality of the voice

can't do it, feel stupid
↳ does attempting in between two worlds

commit – again, voice as something you practice and work on like a guitar, an instrument

my progress, my power-ups

Joe: Do you have any thoughts about why you might find it harder for pitch than for other things, for accent stuff

Carrie: .um yeah, yeah there's physical effort involved in changing pitch so y'know even if I wanted to go up even just a couple of semitones from where my voice is at the moment it requires effort it puts a bit more strain on your vocal folds and y'know trying to be less resonant I mean I am less resonant than I used to be I'm very conscious of that because I keep catching myself going into the like [shift] "the deep man voice" and catching myself and pulling it back up from there like my kids notice it um I don't talk as high - I don't talk as deeply as the other dads um but it's like [pause] [sigh] it's like the difference between changing direction and going up a hill I think you know once you're moving it's easy to change direction but going from flat to going up a hill requires a lot more effort and it sort of feels like that vocally

resonance – perhaps interesting in radio context

effort, strain

conscious of own voice

others conscious of her voice

metaphor

going up a hill

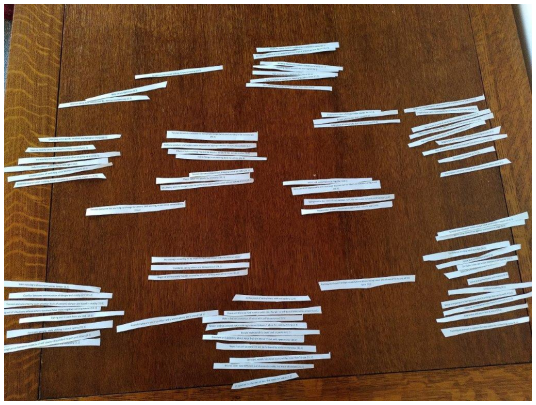
catching myself and pulling it back
↳ in between conscious and unconscious
it's as though it's something bad "catch"

3

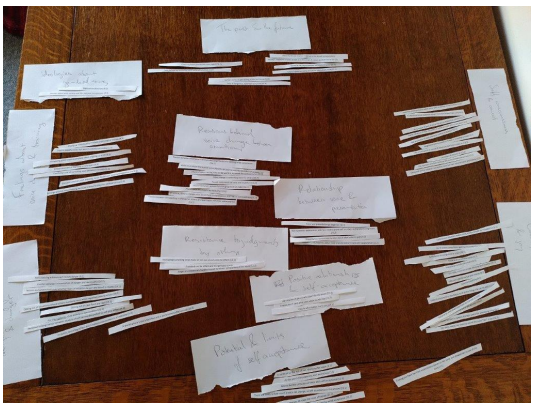
Figure 11.1: Initial noting on a page of Carrie's IPA interview transcript

(a) The experiential statements cut up and placed on a table

(b) Patterns and connections between statements began to emerge as the statements are organised



(c) A structure begins to emerge, with some statements placed in between clusters where they demonstrate relationships between statements



(d) The structure continues to develop and provisional names are given to clusters to describe topics that they relate to

Figure 11.2: Development of a structure connecting Personal Experiential Statements into Personal Experiential Themes

11.1.3.7 Consolidating and naming the Personal Experiential Themes

I then worked back through clusters and returned to working on a computer as I connected the experiential statements back to the initial transcript, finding quotes that supported the statements. At this stage, some experiential statements that were more descriptive and less useful in terms of interpretation were discarded - This is something that Smith, Flowers & Larkin (2022) suggest can form part of the previous step, but I left it until this stage as statements that were dropped did fit into the structure that developed and I found it useful to document this during the process. For example, the theme 'Positive relationships and self-acceptance' only contained three statements and overlapped with themes 'Voice and Presentation' and 'Resistance to judgements by others', so two statements in these theme were re-distributed into these other themes, while the statement 'Admiration of others who don't fit into boxes (23.3)' was dropped at this stage, because this was conveyed quotes supporting other statements. Finally, I organised the Personal Experiential themes into three overarching themes to aid in the presentation of IPA results.

11.1.4 Recorded conversations

Carrie was recorded in two 30-minute conversations in a sound studio with two interlocutors, both cisgender women. The first was Astrid (identified by a pseudonym), a close friend from Denmark who also lived in Glasgow, in the 45-55 age range. The second was Jane Stuart-Smith, an unknown interlocutor, a 58-year-old woman with a near-SSBE accent with some modified regional Northern features. Participants received £10 for each recorded conversation they took part in. They were encouraged to talk freely in the conversations but were given a list of conversational prompts if needed (Appendix C.3).

Conversations were transcribed using Transcriber, then converted to TextGrids, force-aligned, and checked for local background noise and alignment errors following the same procedures set out in Section 9.2.

11.1.4.1 Hand-coding of phonation types

Part II illustrated benefits of prior knowledge of the phonation types used by a speaker for interpreting the results of acoustic analysis of voice quality. In the smaller corpus analysis, this took the form of Phonation Profile Analysis. This auditory-perceptual approach showed both the range of phonation types used by a speaker and the auditory degree they favoured. However, this in-depth analysis had practical limitations in terms of the amount of time needed to conduct it.

In Chapter 9, I instead used an f0-based method to separate creaky voice from non-creaky voice, allowing clearer interpretation of acoustic results. However, this method had significant limitations, as for some speakers no f0 antimode could be identified, and alignment with auditory PPA coding varied considerably between age and gender groups.

Initially, I intended to use the f0-based method to automatically code creaky voice in Carrie's speech. However, after using REAPER to detect GCIs and calculating local f0, as in previous stages of the research, it emerged that there was no clear separation between Carrie's creaky and non-creaky f0. Because of this, I annotated creaky voice manually.

Through the Phonation Profile Analysis conducted in Part II, it emerged that many speakers used one main phonation type as a baseline, with most variation then occurring in different auditory degrees of that phonation type. In approaching acoustic analysis, it is useful to first identify places where a speaker uses a different phonation type, such as creak.

Creaky voice was coded mostly based on auditory quality, with reference to acoustic cues in the waveform and spectrogram. I did not include tense voice in my definition of creak here, and instead coded the types of creak that were included as scalar degrees 2-5 in PPA. For Carrie's voice, low f0 was not a particularly helpful cue for coding creaky voice, as her f0 range in non-creaky voice was wide, and sometimes extended below her f0 in adjacent creak. For example, in Figure 11.3, Carrie's f0 extends down to 61 Hz without taking on an auditorily creaky quality. Furthermore, in Figure 11.4, creak was identified auditorily towards the end of the vowel, and its precise location was identified with reference to dark vertical striations in the spectrogram and a change to the shape of the waveform. However, this shift to creaky voice actually coincided with an increase in f0, not a decrease.

During hand-coding of creaky voice, it also became apparent that Carrie occasionally made use of whisper and forms of harsh voice. Due to these phonation types being those treated as binary rather than scalar degrees in PPA, this meant it was possible to increase the depth of the auditory-perceptual coding of Carrie's voice without requiring the same amount of time as PPA. I therefore coded these phonation types following the same principles as I did to identify these phonation types in PPA as described in Section 6.2.

However, the analysis presented here differed considerably from PPA in two respects. Firstly, I only considered categorical variation in phonation types away from Carrie's baseline quality, that is: use of creak, whisper and harsh voice. I did not attempt to identify whispery voice, breathy voice, modal voice, tense voice, or tense whispery voice, on the basis that variation between each of these phonation types

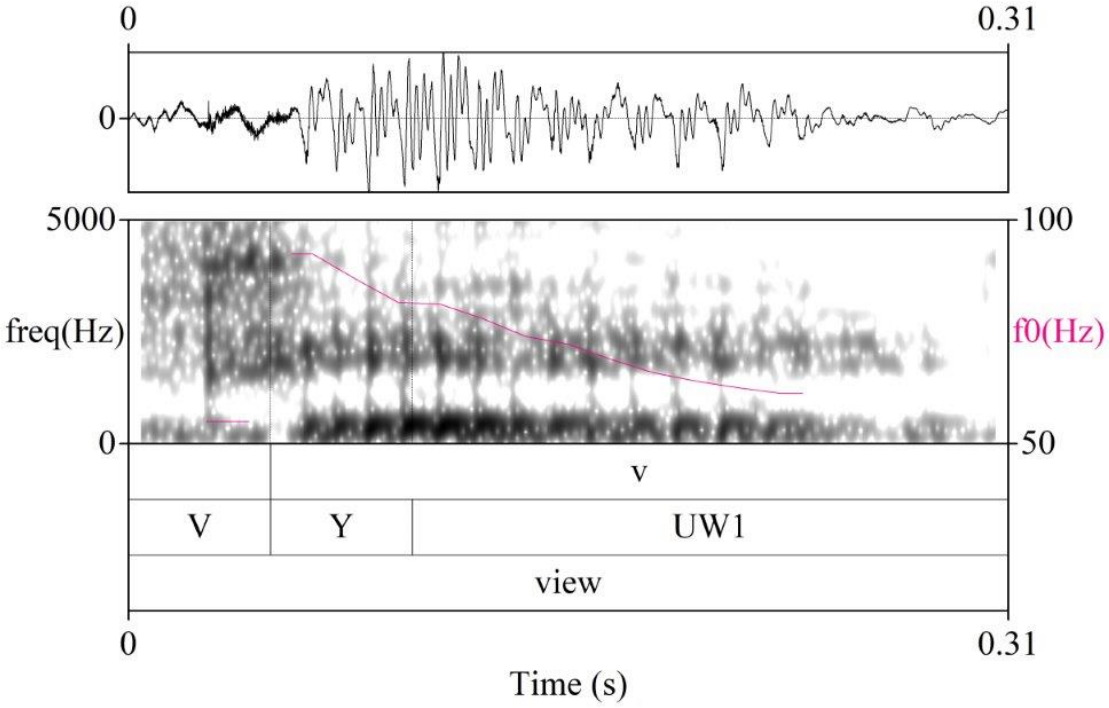


Figure 11.3: Carrie's f0 here goes as low as 61 Hz without shifting to creak

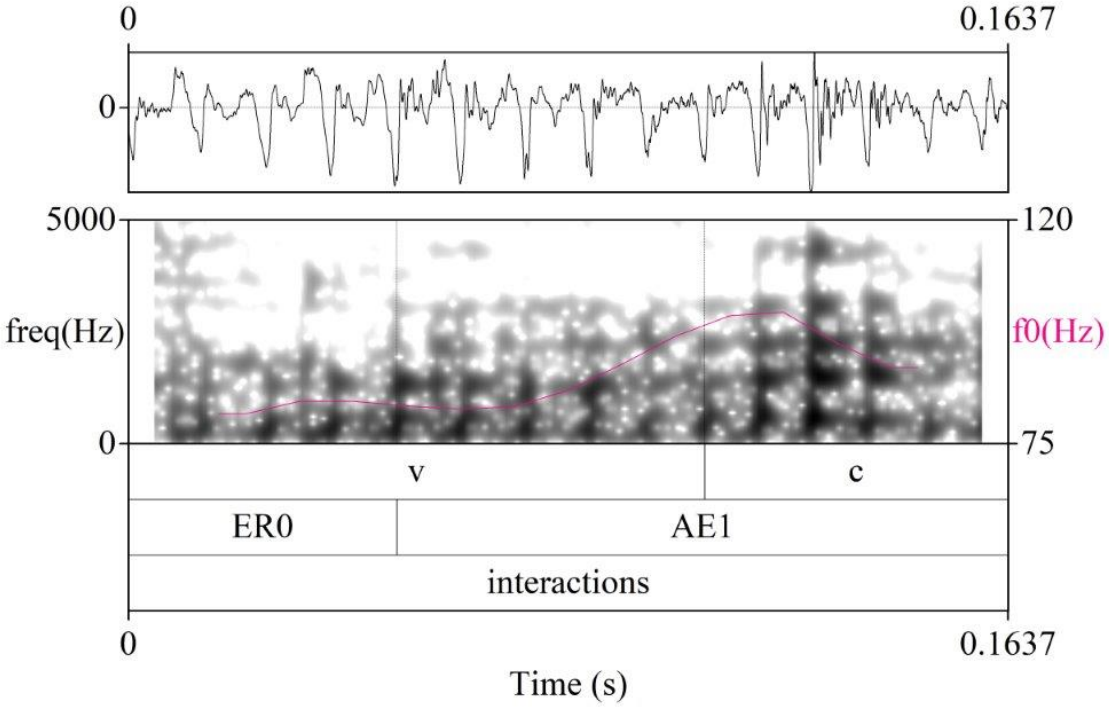


Figure 11.4: Carrie's f0 here is in a creaky stretch (coded as c) ranges from 90-102 Hz, compared to 81-98 Hz in a the adjacent non-creaky stretch (coded as v)

occurs in a more continuous rather than categorical manner, and is therefore better characterised through acoustic analysis. Instead, anything not coded as creak, whisper or harsh voice was considered to form part of Carrie's baseline quality. Furthermore, I did not restrict the length of stretches to 100 ms. This is because this threshold was only necessary to identify phonation variation of a more continuous nature; the start and end points of creak, whisper and harsh voice are usually sudden and can be easily identified through combined analysis of auditory and spectral cues. Furthermore, I allowed Carrie's use of phonation to guide the analysis: For example, allowing consideration of the different types of harsh voice she uses.

Despite these differences, this analysis builds on the same key principle as PPA and the f0-based approach to analysing creak: that first separating different phonation types makes subsequent acoustic measurement results easier to interpret. Furthermore, separating these phonation types allowed closer consideration of times where Carrie uses these phonation types, as described below in Section 11.1.4.2.

An example of what this looked like in practice is given in Figure 11.5

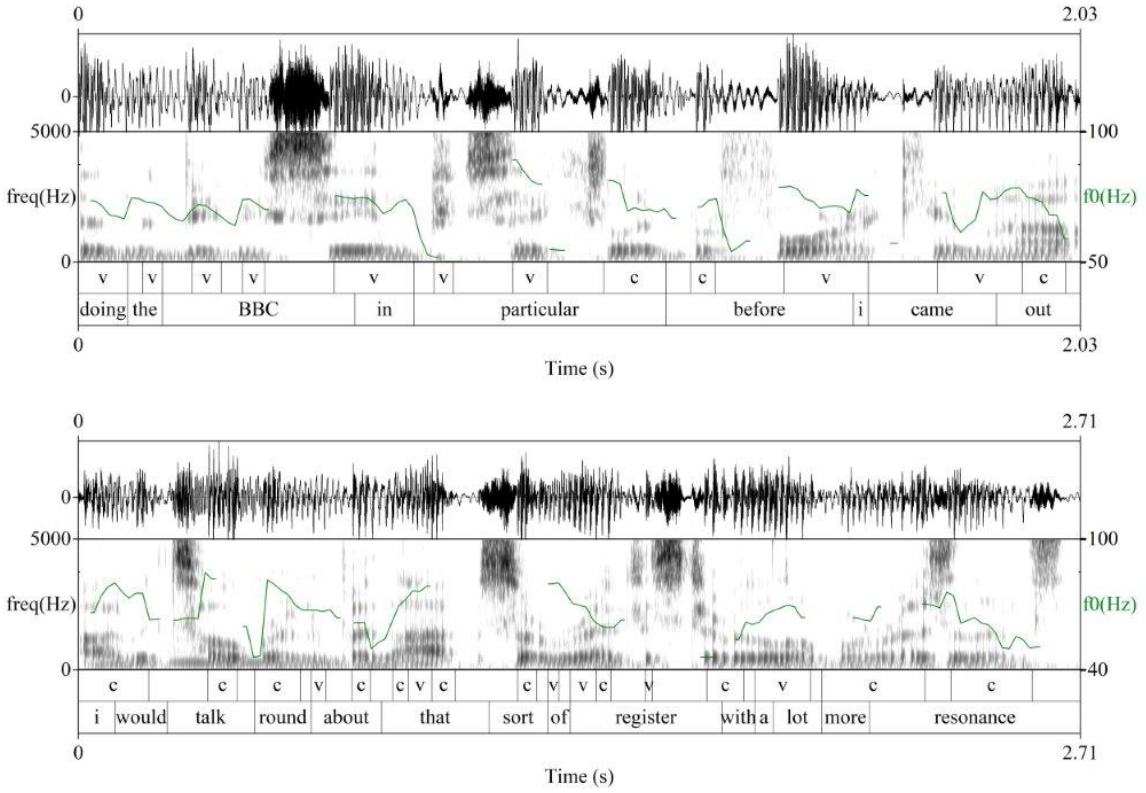


Figure 11.5: Hand-coding of phonation types showing annotation of creak. Taken from Carrie's conversation with Astrid. 'c' indicates creak, while 'v' indicates Carrie's 'baseline'.

11.1.4.2 Integrated qualitative-acoustic analysis of Carrie's 'voices'

The hand-coding of phonation types described in Section 11.1.4.1 allowed shifts to creak, whisper and harsh voice to be considered more closely. These often occurred in constructed dialogue, in instances where Carrie shifted her phonation to convey a particular 'voice' that she used in a certain context, or convey how she believed her voice to be perceived by others. I similarly noted marked shifts in phonation of this kind as they occurred in the IPA interview. The aim of including the shifts in the analysis of the interview was to unify the phonetic and IPA approach, rather than present the two as entirely disjointed, with a hope that this would help analysis of the meaning of Carrie's use of voice quality to be grounded in her own understanding of her experiences with her voice.

The integrated qualitative-acoustic analysis of Carrie's 'voices' therefore involved identifying the names that Carrie gives to each voice, followed by analysis of the instances where Carrie uses a phonation shift to convey these voices, paying attention to the context in which they occur.

11.1.4.3 Acoustic analysis of Carrie's baseline voicing in recorded conversations

Following separation of baseline voicing from creak, harsh voice and whisper, I then proceeded to analyse Carrie's baseline voicing in the recorded conversations. Acoustic analysis of H1*–H2*, H2*–H4*, H4*–2kHz* and CPP was conducted on voiced stretches following the same procedures described in Section 9.2. F0 was also measured in all stretches using STRAIGHT in VoiceSauce.

Overall, the methods of this case study aim to integrate the IPA with phonetic analysis in order to gain insight into how Carrie's voice quality is attributed social meaning. In the next chapter, I begin in Section 12.1 by presenting the results of the Intepretative Phenomenological Analysis of the IPA interview. Then, I consider the results of the integrated qualitative-acoustic analysis of Carrie's voices. Finally, I consider the results of the acoustic analysis of Carrie's voice in the two recorded conversations.

Chapter 12

Results

12.1 Results of Interpretative Phenomenological Analysis

The Interpretative Phenomenological Analysis produced 11 themes, further organised into three overarching themes, which are summarised in Table 12.2. Theme A, explores Carrie's sense of control over her voice and the environment that she uses it in shapes her use of it. Theme B, examines the role of social pressure and fear of other's reactions to her voice how she perceives her own voice and uses it different contexts. Theme C, explores how she portrays self-acceptance of her voice as the way forward.

A The importance of control

A.1 Joy in voice comes from expertise

Carrie speaks positively about her voice from the outset of the interview. Crucially, the love she has for her voice originates from her expert control over it: years of practice have enabled her to use her voice professionally across a range of situations.

Her response to the opening interview question, where I asked her to tell me about her experience with her voice, encapsulates this. She begins by saying, 'I... I love my voice', then lists her professional experience with it:

(1) I started singing in bands [...] then in my mid to late twenties I became a journalist and started getting invited onto the radio and discovered that I could do that and I had a good face and voice for radio and I've been doing that kind of ever since [...] in a typical week I'll do at least one radio thing and in a typical month I'll do a few like other things like get invited on podcasts and things like that so- so yeah my voice is quite important to me

In Quote 1, we see that Carrie's extensive professional use of her voice is central to why it is important to her.

Carrie likens her use of her voice professionally to playing a musical instrument, which suggests both that this is a skill refined through years of practice, and that it is something that she uses in a highly controlled way for creative purposes. We see this here as she describes recording an audiobook:

(2) I was really using my voice I was doing like- you know that thing where you get your voice quieter and quieter, but you go closer and closer to

A. The importance of control	
A.1 Joy in voice comes from expertise	It really did feel like playing guitar but like- I'm not very good at playing guitar, I'm good at- with my voice
A.2 Lack of control over voice and situation creates fear	my voice is gonna be the giveaway and that's when the flaming torches come out and they're gonna run me out of town
A.3 Control over situation brings freedom and joy in voice	'Alright, now I've got your attention! On my terms, y'know, rather than your terms'
B. Feelings around voice training and codeswitching: A tension between self-perception, social pressure, and fear	
B.1 Fear and danger influence use of voice	there's various things that I won't actually do in case the building comes crashing down with me on top of it
B.2 Self-perception of own voice	there's a beauty and the beast thing going on here
B.3 Changing your voice: Something done for yourself, or for others?	A big part of me is now starting think, 'Well, why should I change my voice, just to make you slightly more comfortable?'
B.5 Voice change as physically and emotionally challenging	it was like trying to be a spy, almost, y'know, where you've got your real identity and you've got your secret identity and after a while you're like 'Which one am I supposed to be in this?'
B.5 The relationship between voice and appearance	if I'm not doing the presentation, I don't change my voice
C. Self-acceptance of voice and self as the way forward	
C.1 Self-acceptance and acceptance from others	what do I care what some stranger in the pub thinks of my voice? I- pff- y'know, why am I letting him live in my head, when there's so many more fun and exciting people that can live there instead?
C.2 Potential and limits of self-acceptance	I've gotta deal with the tools that I've got
C.3 The past and the future	I've spent an awful lot of time not doing stuff because I cared too much about what other people thought

Table 12.2: Summary of Personal Experiential Themes, organised into three overarching themes with illustrative quotes for each theme in right-hand column.

> the microphone, so it's almost unbearably intimate and then you drop the punchline! And, so I was doing all that kind of stuff and it- it really did feel like playing guitar but like- I'm not very good at playing guitar, I'm good at- with my voice

She frames herself as being 'good at' her voice here, something repeated several times through the interview. Through this we see that she is highly aware of her skill with her voice, and it is this awareness that brings her joy in her voice:

(3) This is one of things I'm good at, I've spent like twenty something years doing this, um and it was very much the case when I was doing the audiobook as well, it's like- I was loving it.

Her perception of herself as a skilled user of her voice is seen through her comparison of her voice to a tool. Similarly, here we see that her confidence in her voice professionally comes from an awareness of the ways she can use it:

(4) In terms of using it as a tool- I can, I can do lots of things with it [...] I'm very confident now using my voice for that and making what can often be quite a dry subject um more interesting and hopefully funnier as well. And I'm gonna be doing like uh a keynote at [conference name] in a week or so- two weeks and again I'm- I'm very conscious of how I'm gonna use my voice to deliver that cause there's all kind of shifts in tone and there's jokes in there and there's some really serious shit in there and y'know I can deal with that, I think, um and I'm- I'm very confident in terms of what I can do in front of a microphone for music. I- I- I know where my voice can go, what it can't do, why I should never rap again!

Notice here a repeated use of the words 'can' and 'confident' - It is her awareness of what she is able to do with her voice that brings her confidence in using her voice in this way.

Taking joy in using her voice effectively extends to personal interactions. When talking about how she felt her voice differed between the two recorded conversations, she compares her voice to a musical instrument when describing how she uses her voice talking to Astrid:

(5) I was probably using my voice more like a musical instrument actually [...] There's so much to communicate in the way you say the words, rather than just the words.[...] I found I was definitely, definitely doing a lot of that with

Astrid- I do that with my friends cause- there's an element of performance- cause you want to make your friends laugh!

For Carrie, the performance involved in using her voice with friends allows her to enjoy using it.

A.2 Lack of control over voice and situation creates fear

Outside of professional contexts and close friends, Carrie considers herself to have less control over a situation, and in turn this leads to a sense of decreased control over her voice.

(6) What I'm- I'm not so good at is using my voice to go through the world consistently and without fear - yet. So like, I don't, I don't do things like, I don't swim because um, partly because my body shape isn't right yet and partly because I don't want my voice to scare anybody in a changing environment so it's like, I would rather just not go - stay at home.

In these more everyday situations, Carrie's voice becomes something that scares others, and creates fear within herself. This is exemplified in situations where trans women's presence is often called into question and portrayed as dangerous in cultural narratives: in changing rooms and toilets. In these situations, where the reactions of others are unpredictable and outside of her control, she seems to lose a sense of control over her voice, too:

(7) I have this real fear of the man voice being used if somebody tries to strike up a conversation in the ladies' [...] there's still part of me that thinks that - maybe nobody noticed and my voice is gonna be the giveaway and that's when the flaming torches come out and they're gonna run me out of town um so yeah that's a little bit scary

Her use of the passive voice ('fear of the man voice being used') exemplifies this loss of control she feels over her voice. Her use of the term 'the man voice' here suggests a lack of control over the situation too: Here, her voice is gendered not by the gender of its user, but of how others attribute gender to her voice.

She is particularly aware of parts of her voice that she feels are outside of her control in these situations:

(8) That's another thing that I hate is my laugh [...] I've got this big deep 'hurr hurr hurr' laugh and I can't do anything - you can't train yourself - that and coughing - you can't change your cough. I'll be- I'll be out doing the high voice and it's like 'bluerhghhh' it's not fair, y'know. The suffering is real!

Her repeated use of 'can't' here indicates a sense of powerlessness over her laugh and cough, also reflected in her remark that 'it's not fair'. Her hatred of her laugh and cough appears related to a sense that they are out of her control.

Loss of control over her voice brings feelings of dejection and humiliation for Carrie:

(9) Every time that like even one word even comes out slightly lower than you'd intended to, and you're just like 'Right that's it! It's all pointless, nobody's ever gonna accept me as a woman I might as well just go and eat worms!'

The lack of control that Carrie experiences here is not just over her voice, seen here in the 'word' that 'comes out slightly lower' being placed as the subject of the sentence, but in how others will respond to it: Control over her voice is closely bound up in feelings around acceptance from others.

For Carrie, having no control over a situation brings a heightened awareness of her voice. Here she talks about being on public transport with fans of a local football team, a situation that leads her to avoid speaking.

(10) I was standing at the door and they started trying to have a conversation with me from the other end of the carriage and they knew what they were doing! It's like there was absolutely no way anybody in this carriage is gonna hear my voice, because as soon as they do, that's when the shit is gonna go down. [...] It's situations like that, that you're very, very aware of your voice and what connotations it can carry and y'know, that's not a controlled environment like being on a stage or being in a podcast booth

Carrie contrasts her feelings and vocal behaviour in this situation with other more 'controlled' environments. Through not speaking, Carrie opts out of the encounter and of being mocked by others, and reclaims, in a small way, some degree of control over her voice and the wider situation.

A.3 Control over situation brings freedom and joy in voice

In contrast to this, Carrie talks positively about about intentionally using a voice that is contrary to what she believes other people expect from her:

(11) I've deliberately used the man voice while presenting female just to tell people to get the fuck out of my way, um which amuses me immensely because there's nothing remotely tough about me. But I can sound tough, if I want to be, um, and it just throws people!

For Carrie, the voice she's using is 'the man voice' not because of who is using it, but because of what having a 'man voice' represents, which here, is toughness. The way she talks about deliberate use of 'the man voice' here comes in stark contrast to how she discussed it in the context of public toilets and changing rooms, where she frames its use as outside of her control.

Deliberate use of this voice in particular situations where she is in control of her voice and is using it for a specific purpose is positive, and this is exemplified in how she uses her voice on stage:

(12) When I started doing open mic nights I was wearing a dress on stage, I was quite clearly presenting female and I think people were expecting some kind of falsetto musical theatre thing and I came out sounding like Lemmy from Motörhead, which was fun actually that was- that was fun, because it was like 'Alright, now I've got your attention! On my terms, y'know, rather than your terms'.

Carrie realises that people have expectations about what her voice will sound like and that her singing voice, which she likens to a male singer, goes against these expectations. The control she has over her voice and the situation ('on my terms') in these contexts is what makes using her voice in this way 'fun' and 'amusing', turning the reaction and attention from others that she gets when using a masculine vocal style into something positive.

Fearing negative reactions from others in uncontrolled environments heightens her enjoyment in creating a deliberate disconnect between her voice and her gender presentation. She summarises this here:

(13) That's why I take so much joy in it on stage because I can control it, I can decide what voice I'm gonna use and how I'm gonna project it and use it to shout over the noise of everybody in the room

B Feelings around voice training and codeswitching: A tension between self-perception, social pressure, and fear

B.1 Fear and danger influence use of voice

Carrie's fears around how others will react to her voice are a major factor in how she uses her voice in her everyday life.

She demonstrates how instrumental her voice is in how others treat her in the following example about how she was gendered by staff at a gig:

(14) I think voice is a safety thing as [...] the waiting staff saw me before they heard me speak and every single one of them addressed me as female. And the security guys at the gig heard me before they saw me and sir'ed the absolute shit out of me, even though I had the wig and the makeup and all the rest of it. So there- there is an element of feeling that you have to change the voice in public places or more often than not just not talk.

In Carrie's experience, voice takes precedence over appearance in how others treat her. This leads her to view ensuring that others do not notice a disconnect between her appearance and her voice as a necessity ('you have to change the voice'). This affects how she uses her voice in public places, often choosing to avoid talking to avoid danger.

While Carrie draws on silence to move around public spaces unnoticed, she associates speaking with danger:

(15) Nobody's really even noticed I was there until they heard me speak and th- head goes round like that and then I start like- it's like that meme of the Simpsons like 'I'm in danger' [See Fig. 12.1] [...] I've learnt the hard way that men are really, really creepy to any feminine presenting person in the world but this is different to that it's more aggressive than that and I don't like it, um women don't react in the same way

She reports that there is a gendered aspect to how others react to her presence, with men reacting aggressively. She talks about perceived incongruence between her voice and appearance causing discomfort not only to others, but also to herself:

(16) It feels really strange having a different voice to your presentation. So if I'm going out in boy mode, which I- I still do, because y'know, I don't want to

Figure 12.1: A popular meme that Carrie references to convey her feelings of danger (*Ralph in danger* 2014)

> attract attention like if I'm with my kids or whatever, um, I- I- I couldn't use that learned voice with it.

In turn, this 'strange' feeling that accompanies having a different voice to her presentation affects how she uses her voice in everyday life.

Carrie's decisions about whether to present as female in a certain situation, and in turn, what voice to use, are also closely connected to the perceived severity of consequences in a certain situation. This is exemplified by how she opts out of presenting female with her children:

(17) I have a limited amount of time with my children and I don't want somebody to mess that up by taking exception to me [. . .] it's much, much easier just not to play that game

This decision to opt out of situations where the reactions of others have the most negative consequences mirrors her decision to avoid talking in certain contexts.

For Carrie, her decisions regarding what voice to use in a situation differ from regular style-shifting:

(18) We all change our voices according to who, who we're with and what we're doing, all of us, and for trans people I think [...] it's a bit more serious than that in public places with strangers. I think that's the main difference: It's not to try and fit in with our peers it's to hopefully not get singled out and murdered

Here, she contrasts what she sees as the usual reason for style-shifting ('to try and fit in with our peers') with what she sees as the main reason that trans people style-shift, for reasons of safety.

Despite identifying safety as the main reason for style-shifting for trans people, there is tension between her experience with feeling fear around others' reactions to her voice, and a tendency to minimise that fear. Directly following her observation that trans people style-shift to avoid getting 'singled out and murdered', she says:

(19) I am deliberately exaggerating for comic effect but for a lot of people in a lot of places that's not a joke, that- genuinely to be trans is to be marked out and be in dangerous situations

She feels a need to say that she is 'exaggerating' the danger, but then goes back on it. She reflects on her own uncertainty about whether her fears are based in reality:

(20) There's an aggression that I don't want to see when I'm with my kids. Whether it's- whether it's there or whether I'm imagining it, I don't know - but I'd just rather not have to work that out

Her descriptions of times where she has felt in danger (e.g. 'there's an aggression', 'it's like that meme of the Simpsons like "I'm in danger"') are contrasted with a use of humour to minimise a sense of danger (e.g. 'I am deliberately exaggerating for comic effect').

She reflects on the conflict she experiences about the reality of her sense of fear and how it affects her behaviour in the world here:

(21) There's one world where I'm proud of who I am and there's one world where I'm terrified that everybody's gonna kill me - and there's no evidence that that world is there, other than years of being in the closet and being scared to come out, because that's what I thought the world was gonna be like. And I think a lot of this is to do with that, it's not necessarily that it is a problem with many people, but I've convinced myself that the risk is probably higher than it actually is. In much the same way that I'm scared of heights and there's various things that I won't actually do in case the building comes crashing down with me on top of it, it's probably not gonna happen, but y'know, I'm not going up there.

Carrie see her own fears as being irrational, despite also giving examples of times that others have directed mockery and aggression towards her.

B.2 Self-perception of own voice

Carrie talks negatively about her own voice at times and her own negative perception of her voice is tied up in how other people perceive and react to it:

(22) I was at a open mic thing recently and talked to some cisgender folk afterwards and they were like 'I didn't realise you were trans until you opened your mouth' and it's like- and they did an impression of me which was like '[harsh voice]' it's like 'Yeah okay, thanks dudes'. They didn't mean it in a nasty way, people rarely do, um but it's, it kind of reinforced what I- what I think and I- I still think I've got a great voice, but I just wish it wasn't... a man voice

Her voice shifts when she voices how she believes others to see her voice, here shifting into harsh voice. She refers to her voice here as 'a man voice', which here seems connected again to how others attribute gender to her voice.

She also uses harsh voice to convey how others perceive her voice here, where she compares her voice to that of a cisgender friend that she works with on the radio:

(23) She's a cisgender woman, her pitch is much higher than mine, she speaks naturally in a way that I don't [...] I do sort of feel that there's a beauty and the beast thing going on here! Where like [name] will say her thing and then I'll go [harsh voice] '*aurgh 'ello!*'

Carrie compares not only the features of their voices here ('her pitch is much higher'), but frames her friend's use of her voice pitch as natural, which in turn positions her own voice as unnatural. This is furthered by her description that there's 'a beauty and the beast thing going on'.

Her self-perception of her voice is conveyed through a decrease in pitch in the following example, when she mentions 'the deep man voice':

(24) I'm very conscious of that because I keep catching myself going into the like [lower pitch] '*the deep man voice*' and catching myself and pulling it back up

I consider the phonetic form of these shifts in the integrated qualitative-phonetic analysis in Section 12.2.

B.3 Changing your voice: Something done for yourself, or for others?

Carrie often remarks that she is very conscious of her voice, which seems heightened in recording contexts:

(25) But in the studio yeah, you hear your own voice, and I tend to turn things up quite loud. So, I'm very aware of the audible breaths that I have between sentences [...] that's probably when I do my best voice I think, because I'm conscious of it

This self-awareness of her own voice in the studio, where she hears her own voice played back to her, is an important factor in her producing a particular vocal style, which she variably refers to as her 'best voice', her 'radio voice' or her 'work voice'.

At times, Carrie's motivation for the vocal style she adopts on radio appears to be something that she does for herself:

(26) I want my voice to be the most 'me' it can be and because I'm, I'm hearing it so loud [...] I- I try that wee bit harder, I think

Here, her adoption of this style is something that she does to make her voice sound like her own as she hears it played back to herself.

She discusses enjoying using this voice because of how she is able to use it to create intimacy with listeners:

(27) I try to raise my voice and do more variety and talk much more softly and it's- it's something I think that- that being on a mic with headphones encourages you to do anyway. I've never really been one of these people that sits way back from the microphone and just booms into it, I'd much rather be like close and quiet, cause I know, cause radio's like really intimate, and I really like that. The thing I've always loved about like radio or podcasts or audiobooks is it [whisper] *feels like they're talking just to* me, y'know, they're [whisper] *just* there telling me about this amazing thing and I want to try and do that.

While she talks positively about using this vocal style because of the effect it has on listeners, there is a tension between this and another motivation: To appease others and soften their reactions to her trans identity:

(28) There's also the knowledge that- that when I'm on like [major radio station] there's an [whisper] *awful lot of bigots who are listening to it.* And I don't really want to give them an opportunity to go 'oahh'.

This sentiment is also echoed in her descriptions of her 'phone voice':

(29) I got a phone call from [radio station] earlier and I immediately raised the pitch of my voice to do a much higher 'hello' than if it had been somebody I knew [...] and even with that there was a pause when they say 'Hello, can I speak to Carrie?' and I go [pitch increase] *'Hi! Yeah, I'm Carrie!'.* 'Right, pfff', as they go- you can hear the wheels turning in their head. It's like 'Hang on a minute!'. Um, so anything I can do to sort of reduce the duration of that intense awkwardness, I do.

Here, the motivation for increasing the pitch of her voice is to reduce the 'awkwardness' - to conform to other people's expectations of what a woman's voice should sound like.

In contrast to the positive way she talks about radio voice, her use of phone voice is much more begrudging:

(30) A big part of me is now starting think, 'Well, why should I change my voice, just to make you slightly more comfortable?'. So on the one hand I will do it on the phone, in the hope of minimising the awkwardness, but a lot of times I'm just like 'Y'know, I look like this and I sound like this and I've got a girl's name! So what!"

There is a tension here between a resistance to changing her voice for the benefit of others, but doing so anyway to minimise awkwardness.

This feeling that voice change is something done for others is clear where she talks about the role of changing the pitch of your voice for accessing transition healthcare:

(31) Until very, very recently, the only get any sort of gender related healthcare was to fit every stereotype [...] Transition was often determined on the basis of whether the guy thought he would want to shag you afterwards [...] We were just basically told, 'Right, if you want to be able to go and transition from, y'know, male to female, for example, you need to get your legs out, you need to look conventionally pretty, you need to lose weight, you need to have a high pitch voice!'. Whether your physical structure was capable of doing that voice - you still had to try for it, because if you didn't, then you wouldn't get stuff.

Carrie places the need to have a high pitched voice alongside other stereotypes of womanhood that trans women are often required to conform to in order to access healthcare, also listing the need to wear a skirt, makeup and a wig.

When Carrie considers whether she may return to voice therapy in the future, she compares the voice to that of 'a children's TV presenter':

(32) I might revisit the voice thing [...] Go back to like talking like a children's TV presenter, but I don't know, I don't know if I can be arsed!

This comparison suggests that the voice that she was learning in voice therapy was a particular vocal style that does not necessarily suit her and her life: she is unsure if the effort of voice therapy is worth it to end up sounding like someone else.

B.4 Voice change as physically and emotionally challenging

Carrie talks about her experiences with voice training and changing the way she uses her voice between different situations as both physically straining and emotionally challenging.

Though she talks positively about the speech therapist she saw for voice therapy, she has mixed feelings about its effectiveness because it did not allow for codeswitching:

(33) It was fine and it was working but I co-parent, so I'm with my kids half the week and I found that I just there's a codeswitching that happens

Her opinion on voice training here seems very neutral ('it was fine'), and her remark that 'it was working' comes in contrast to examples of times where it did not seem to 'work', and her discussion of how it became difficult to manage using different voices in different situations:

(34) There was times I'd be presenting as me, and I'd, y'know, forget to use the voice and I'd be really embarrassed and it was just- it was- it was too much. It was like- it was like trying to be a spy, almost, y'know, where you've got your real identity and you've got your secret identity and after a while you're like 'Which one am I supposed to be in this?'

This appears to have an emotional cost: using a voice that she sees as not matching her presentation brings feelings of embarrassment, but also of secrecy. It is unclear

which self and which voice she sees as being her 'real' and 'secret' identities, but it seems that feeling she has to change her voice and forgetting to do so heightens feelings of inauthenticity that she feels in having to change her presentation depending on context.

Her ability to produce and sustain a higher voice depends on how appropriate she feels it is to her wider style and the wider context:

(35) I can do it when I sing! No problem at all. I mean, I've got a very similar vocal range to like Thom Yorke I can do all that falsetto stuff, no bother! But I just felt like an idiot, y'know, and I've described it- the best way I can describe it is trying to be a children's TV presenter, where you've got your eyes open wide and like a really mobile face and everything thing is [shift] *moving up and down and you're talking like that.* It's like, I can't do it. I just feel stupid.

Though she is able to produce a higher pitch in singing, the vocal style that she was learning in voice therapy appears to involve more components, involving shifts in intonation and embodying a different persona: the children's TV presenter. She can produce this voice, as she imitates it in the moment, but also believes that she 'can't do it', suggesting that the issue is connected to whether the voice feels authentic to her, and to sustaining it in everyday context while maintaining the ability to style-shift. Trying to do this is and not feeling able to seems emotionally taxing ('I just feel stupid'), which she elaborates on as she continues:

(36) As long as I'm straddling two worlds, it's not gonna work, 'cause I just can't commit to that muscle memory thing. It's like- yeah, muscle memory's quite a good way to put it, it's like if you play guitar right-handed all the time and then for a couple of weeks you just switch to left-handed for no reason, and then you pick it right again, and it's like, 'How do I do this again?' I've lost all my progress, and I've lost all my power-ups, I think, with my- with my voice as a result of that.

Continuing to attempt to shift her voice between contexts while learning a new voice results in Carrie feeling dejected, but is also physically challenging:

(37) Joe: Do you have any thoughts about why you might find it harder for pitch than for other things, for accent stuff?

Carrie: Um, yeah, yeah- there's physical effort involved in changing pitch, so y'know, even if I wanted to go up even just a couple of semitones from

where my voice is at the moment it requires effort it puts a bit more strain on your vocal folds [...] It's like the difference between changing direction and going up a hill, I think, you know once you're moving it's easy to change direction but going from flat to going up a hill requires a lot more effort, and it sort of feels like that vocally

B.5 The relationship between voice and appearance

As we have touched on already, Carrie does not present as female consistently in all aspects of her life and her gender presentation is an important factor in her vocal behaviour:

(38) if I'm not doing the presentation, I don't change my voice, and you know and I don't, I don't go stand with the dads either - cause I've got nothing in common with the dads - um, and I'm not moderating my voice to try and sound more masculine and I mean y'know I'll turn up in like my raspberry coloured coats and y'know fairly feminine colours and stuff

Here, when she talks about not presenting female around her kids at school, she is not describing 'presenting male' necessarily - other parents are aware of her trans identity and she is not making an effort to 'sound more masculine' or avoid dressing in a feminine way.

How well other people know her and if they know that she is trans appears to be an important factor in how the relationship between voice and presentation functions:

(39) If I'm going to the chemist to get my prescription [...] I've not got my wig on and I've not got my makeup on, it's like 'Ugh am I gonna do the voice? Uhh, I don't know' and these days I just don't, 'cause they know me now, but it's like in- in the early days I would make a point of- of getting all scrubbed up like this and doing the high voice

Talking about how she uses her voice with Astrid, a close friend, she says:

(40) Astrid's seen me in boy mode hundreds of times, um, and I know she doesn't care, and I kind of feel like that with my voice as well: It's like, we're friends- don't think anything less of me if one day I sound like Barry White and another days I sound like Kylie Minogue!

Here, her close friendship with Astrid is important in both how she presents around Astrid and what she does with her voice - Knowing that Astrid accepts her and 'doesn't care' brings her freedom in what she does with both her voice and her presentation.

She speaks about her acceptance of her voice paralleling a similar journey in her presentation:

(41) So many of us are gonna go, 'Y'know what, no, I'm not doing it! Just because you think this is how a woman should talk - I'm not gonna talk like that because I am still me'. And it's like, I found that- that- that- this is something that happened with my physical presentation which is that when I first came out, I had to basically - I wasn't quite dressing like a barbie doll [...] but I felt that I needed to prove to everybody else that I wasn't a guy and the easiest way to do that is to go hyper-femme

We see here a resistance to conforming to sexist standards placed on her voice ('I'm not gonna talk like that') and a sense that this vocal style would be inauthentic ('because I am still me'). She frames this journey of resisting the expectations of others as paralleling a journey in her presentation: She no longer feels the need for her voice or her presentation to conform to standards of femininity set by others in order to 'prove' her gender to others.

Elsewhere, though, positive feelings that she has around her voice are contrasted with the way she feels about her appearance:

(42) I'm very aware of, of not looking the way that I want to look, without wearing a wig and wearing makeup and all the rest of it, so I don't think, um, that's ever gonna change. But with the voice thing, there's definitely been a change, there's definitely been a switch and it started with the music where I was just like 'Uh! This is what I sound like!' And y'know- y'know- I'm a trans woman singing in a rock band, that's what we sound like! Y'know, um, and it's a good thing.

The contrast here comes not only in how she feels about her voice ('it's a good thing') and her appearance ('not looking the way I want to look'), but also in terms of whether it is possible for those feelings to change.

She discusses this feeling of immutability elsewhere, relating it to the age that she began her transition:

(43) I've described this before as I kind of feel like my voice is a consolation prize cause, y'know, I didn't put it all together until my forties, so I can't ungrow

> all the things that testosterone did to me and I can't make my vocal chords magically change and I wouldn't want to risk any sort of surgery because my voice is just such a crucial part of what I do, um, but I- I think I've got quite a good singing voice, particularly now I've had years of practice with it

There are mixed emotions in this description of her voice. On the one hand, her voice is framed as a prize and she describes it 'a good singing voice'. However, positive feelings around her voice come only as a 'consolation prize' - at the expense of feeling a sense of having lost in terms her physical appearance ('I can't ungrow all the things that testosterone did to me'). Her remark that 'I can't make my vocal chords magically change' also suggests that, were this possible, she would make her vocal chords magically change, making this 'consolation prize' feel bittersweet.

C Self-acceptance of voice and self as the way forward

C.1 Self-acceptance and acceptance from others

Carrie talks about becoming increasingly resistant to feeling judged by other people:

(44) It's not a 'me' problem it's a 'them' problem, and I'm increasingly unwilling to put up with that shit!

She re-frames the issue here, making it no longer a problem with her voice but a problem with how others treat her because of it.

Carrie relays a narrative about attempting to change the car registered on her insurance over the phone, where her interlocutor would not accept that Carrie was the caller because of the sound of her voice. Telling this story, she used increasingly emotive language:

(45) It really bugs me, because it's like, why does what genitals I may or may not have dictate whether you can change the fucking registration number on my car insurance!

She becomes increasingly frustrated at the interlocutor's refusal to take her for who she is, and creating unnecessary hurdles in her day-to-day life. This anger is directed towards other people, rather than toward herself and her own voice.

A realisation that the standards that are set for her voice by others are disingenuous appears to be involved in her resistance to the judgements of others:

(46) Like if I have the most feminine voice in the world they're still gonna be like 'Ah, but look at your shoulders, eh you're not really a woman are you?'. Y'know and it's like that: There's always gonna be something else that excludes you, so why- why chase after the approval of people that are never gonna grant it?

Resistance to these standards also seems to come from positive relationships in Carrie's life:

(47) Why chase after the approval of people that are never gonna grant it? And y'know there- there's- because you know the people that actually matter don't care, they really don't! And the people that do care are not worth your time. I think, y'know, so it's like what do I care what some stranger in the pub thinks of my voice? I- pff- y'know, why am I letting him live in my head, when there's so many more fun and exciting people that can live there instead?

The idea that people who matter to Carrie do not hold her to these same standards allows her to resist against the pressure that she feels from others for her voice to sound a particular way.

C.2 Potential and limits of self-acceptance

Carrie speaks about accepting her voice, but this comes with limits and exceptions:

(48) I'm kind of at the point now where, uh, just sort of accepting it rather than being particularly bothered by it the exception is [...] if I get a cold or I've been doing a lot of singing, my voice goes lower for several days afterwards and I'm really self-conscious about that

The acceptance that she has for her voice is not consistent with her feeling very self-conscious at times where her voice is lower.

Times where she speaks negatively about her voice, though, are usually paired with a positive statements about its potential. This can be seen in response to a question about whether she feels that her voice fits her:

(49) I think it's a terrible flaw! I would– I would like to have Taylor Swift's voice [. . .] but y'know, I sound like Lemmy from Motörhead, or Kurt Cobain out of Nirvana or any of these guys, so I'd– I've gotta deal with the tools that I've got.

Despite referring to her voice as a 'flaw', the comparisons that she makes between her voice and those of popular male singers are not negative ones - elsewhere she talks positively about the effect that sounding like Lemmy has on audiences. Instead, she positions her voice as a tool, saying 'I've gotta deal with the tools that I've got'.

Her desire to accept her voice seems to stem from seeing limits to the flexibility of her voice:

(50) You're not gonna go, 'Fucking hell it sounds like Taylor Swift!' Y'know what I mean? It's like, it's still gonna sound like, like me, um, so I- I- I wonder- I wonder how much the answer to all this stuff actually is - self-acceptance and hoping that social acceptance will follow, like if we just stop giving a shit maybe other people will stop trying to make us change.

Carrie sees self-acceptance as a way forward here, and as a way to resist the standards placed onto her voice and the voices of trans women more generally.

C.3 The past and the future

As in the previous extract, Carrie often presents optimistic visions of the future, where self-acceptance plays a pivotal role, both in her own relationship with her voice and more generally.

This optimism for the future is presented in the context of and in contrast with descriptions of the past. Here, the past is defined by caring too much what others think:

(51) There's a realisation that I've spent an awful lot of time not doing stuff because I cared too much about what other people thought, when those people didn't care at all about me and it's like well, why?

Carrie talks about the past as dangerous for trans people, but positions this next to optimistic sentiments about the future:

(52) You couldn't come out! It was too scary, too dangerous, um- and I think- I think, where we're headed now, despite everything that's going on, is a much better place, where it's just like- like my kids just don't give a shit, their peers just don't give a shit!

When she looks forward to the future, not caring about what others think as an important part of her hope for the future:

(53) I'm quite looking forward to the stage of [...] Y'know, like, grannies who just don't give a fuck, right where it's just like, 'Yeah, I'm gonna have my hair blue and I'm gonna wear all these colours, and I'm gonna toot about and just get the fuck out of my way, I've got stuff to do!' Right, they just- no shits are given anymore!

Carrie seems to already see herself as already caring less about the judgments of others, but continues to look forward to being able to care less about this as she ages. Ageing is also a central part of this process for Carrie, appearing to free her more from fears about how others see her. Here this vision is presented in terms of other aspects of her presentation, but elsewhere, the voice takes on a more central role. In this extract, she talks about how her admiration for the comedian Suzy Izzard,[1] and about her desire to continue to become more confident in her voice:

(54) Eddie came on stage dressed as a woman, spoke like a guy and all the cool girls were so into him! It was like- that's a vision that I could get behind, because I would like to be that person that- is wearing what they want to wear, presenting how they want to present, talking just as they talk and everybody going 'Hey! You're awesome!' [...] I would really love to have that confidence to just go, 'Well, y'know, this is my voice' and if you don't like it, there are other voices you can go and listen to instead, but just not mine.

This vision of having more confidence in her voice does seem in part connected to others having positive opinions ('all the cool girls were so into him!'), but in part to a refusal to change in cases where this is not the case ('there are other voices you can go and listen to instead').

This positive vision of the future that involves confidence in her voice extends to trans women more generally. When talking about the pressure to present and speak a certain way in order to access transition healthcare, she says:

(55) There's so many more of us are feeling confident enough to come out and to be ourselves and live our best lives without- y'know, without apology or fear or anything else

This optimism exists also for how society and the healthcare system treat trans people more generally. Towards the end of the interview, Carrie talks about accessing

[1] Izzard has specified she that uses she/her and he/him pronouns and continues to be known by both Suzy and Eddie(Izzard 2023)

transition healthcare due to long waiting lists for NHS gender identity clinics and underfunding, but the frustration that she expresses about this issue is also paired with optimism about the future:

(56) It's just really, really frustrating at the moment, because just so much of the oxygen is being taken up by just idiots talking nonsense [...] but- y'know this is- y'know this is something that is very much of a time and of a place and I don't think it's gonna be a long term thing. I think in the long term, we're gonna we're just gonna understand that some people are trans and some people will want healthcare and if we do that it makes them live happier lives and this is what healthcare is supposed to be about.

Carrie appears here to be at a midpoint between a past that she sees as scary and stifling, and where the judgements of others impact how she lived her life, and an ultimately more positive future for herself, where confidence and self-acceptance play a role in a more positive future for trans people and the next generation.

12.2 Integrated qualitative-auditory-acoustic analysis of Carrie's 'voices'

The IPA revealed a number of potential influences on Carrie's voice. For example, control emerged as an important aspect, with Carrie discussing using her voice in different ways when using it in controlled contexts (on the radio, for an audiobook, and in music) and in situations where she felt out that others reactions to her voice were out of her control and where she feared for her safety. Throughout the interview, she also identified a number of different 'voices', and on occasion produced examples of these in both the IPA interview and the recorded conversations. In this section, I outline how she names and describes these different voices, and provide a qualitative acoustic analysis of examples of each of these where she produces them. Finally, I consider a few examples of other instances where Carrie shifts away from her baseline voice and consider why this might be occurring.

She identifies several different voices at one point in the interview, which I take as the starting point for this analysis:

(57) There's work voice, there's friends voice, and there's all-the-people-that-are-gonna-misgender-me voice

Here I look at how she consider Carrie's 'work voice', 'friends voice' and 'all-the-people-that-are-gonna-misgender-me voice', along with other voices that emerged in the IPA interview and recorded conversations: the 'man voice', her singing voice, 'beast' voice and staying silent.

12.2.1 'Work voice'

Carrie refers to her 'work voice' to cover both her 'radio voice', which she uses on air, as well other ways she uses her voice professionally, such as in recording an audiobook. She described this voice in the IPA interview, describing some of the changes that this involves in Quote 2 and Quote 27, and talks about some of the motivations behind these changes in Quote 28 and Quote 28. To recap, she described 'work voice' as involving raising her voice, increasing 'variety' and talking 'much more softly'. It often involves speaking quietly while approaching the microphone to create intimacy with the listener, trying to make them laugh, and avoid negative reactions from the 'awful lot of bigots who are listening to it'.

She gives several examples of the voice she uses on radio. While two of these are expressly noted as being her 'radio voice', I also examine times that she appears to use

whispery phonation in a way that mimics how she shifts her use of phonation on radio as she discusses her radio voice.

In the recorded conversation with Astrid, she gives an example of how she used her voice on the radio before she came out. Shown in Figure 12.2, her pitch is very low here. As we will see in Section 12.3, Carrie's median f0 is 103 Hz in baseline voicing in the rest of the conversation with Astrid, compared to a median of 72 Hz within this stretch in non-creaky stretches. But more is going on in this shift than a drop in pitch alone: as the utterance progresses, Carrie uses more creak, with her f0 in these creaky portions dropping as low as 35 Hz (measured manually).

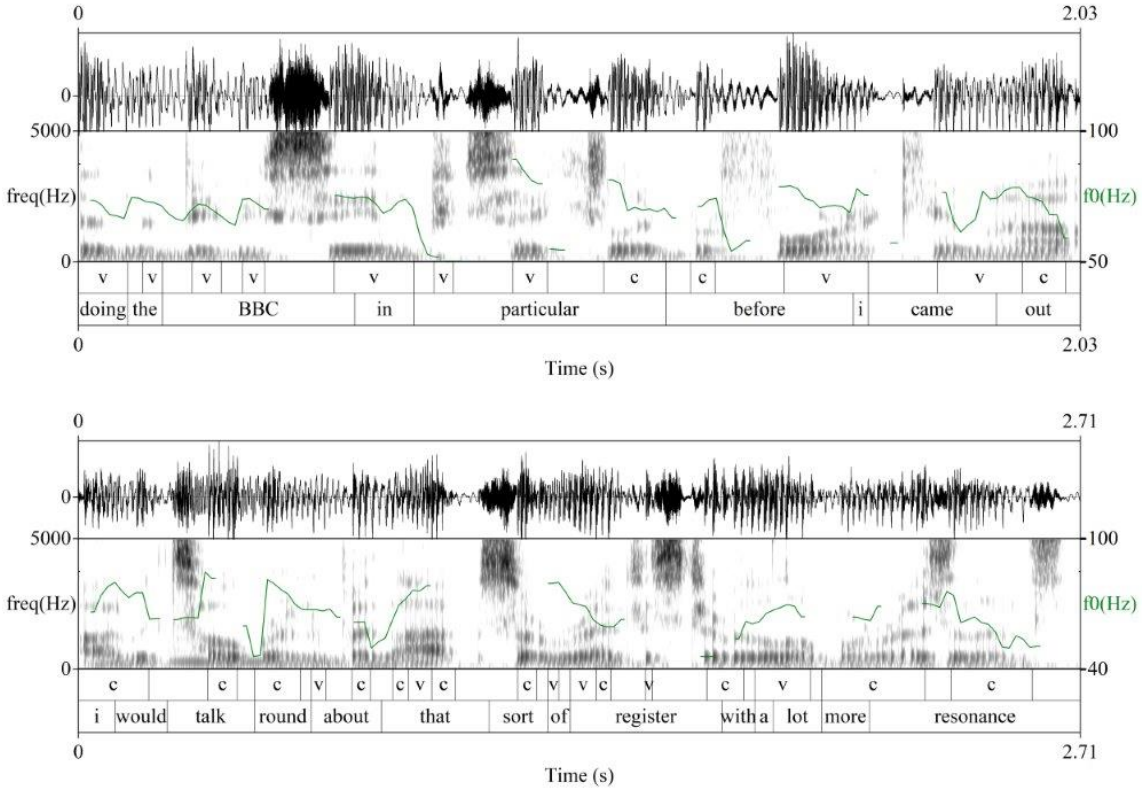


Figure 12.2: Wideband spectrogram showing Carrie's impression of her 'radio voice' from before she came out. Taken from Carrie's conversation with Astrid.

She produces a similar drop in pitch in the example in Figure 12.3 from the interview when she discusses how she codeswitches between her different voices. However, this shift is marked not by the presence of creak, but instead by a breathy quality. This can be seen in the spectral slices in Figure 12.4, both taken from /o/ vowels. While /o/ from 'codeswitch', before the shift into Carrie's radio voice begins, shows a negative H1-H2 and flat slope until 2 kHz, the /o/ of radio shows a steeply positive H1-H2 and overall slope until 2 kHz.

Carrie also emulates another aspect of her radio voice at another point in the interview, where she discusses how she enjoys the way she can use her voice to create intimacy with the listener. As shown in the narrowband spectrogram in Figure 12.5, she shifts into a whisper here, recreating this effect of intimacy.

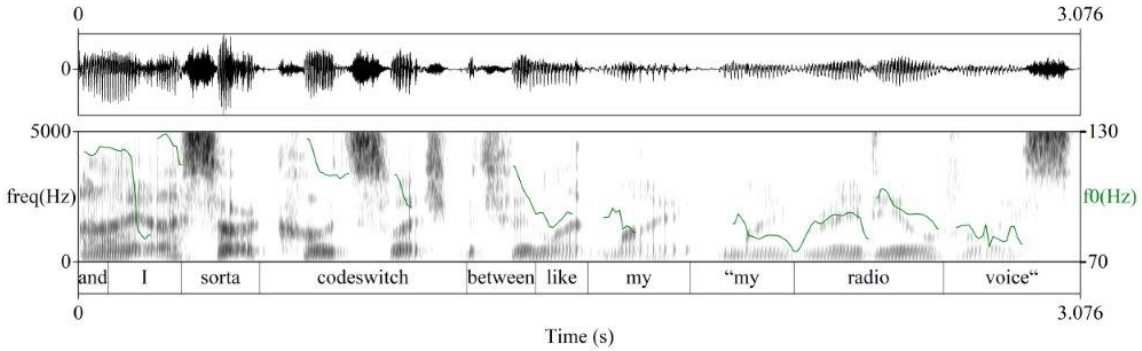
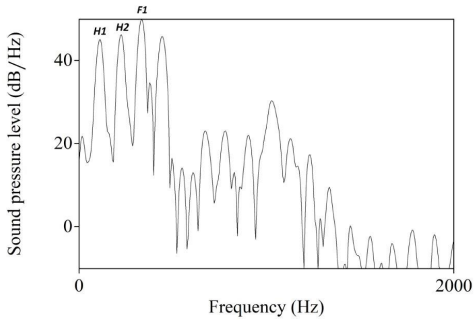
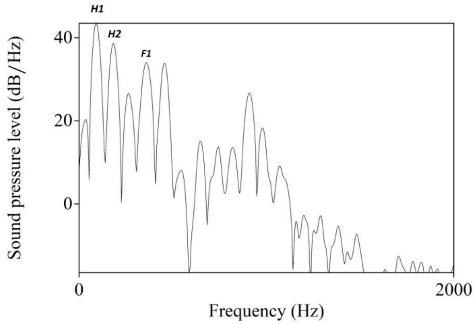


Figure 12.3: Wideband spectrogram showing Carrie's reproduction of her 'radio voice' in the interview.



(a) /o/ of 'codeswitch' in Figure 12.3. F1 is visible at 334 Hz. Taken from Carrie's interview.



(b) Spectral slice taken from /o/ of 'radio' from the same production as Figure 12.3. F1 is visible at 358 Hz. Taken from Carrie's interview.

Figure 12.4: Spectral slice taking from /o/, comparing Carrie's radio voice to a nearby token of the /o/

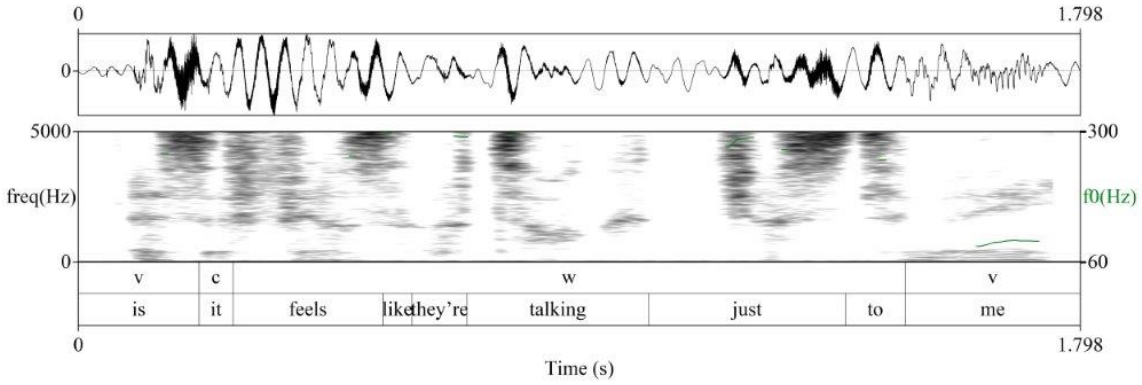


Figure 12.5: A narrowband spectrogram showing whisper, annotated as 'w'. Voicing visible in pitch contour, annotated in green, towards the end of the utterance.

12.2.2 'Friends voice'

Carrie discusses how she uses her voice around friends several times in the interview, particularly in response to a question about how she felt her voice changed between the two recorded conversations she had.

'Friends voice' involves two main components. On the one hand, as shown in Quote 40, it seems that Carrie knows that her friends accept her, and so she feels no pressure to produce the 'learned voice', and because of this can incorporate a wider range of gendered styles into her speech - ranging from 'Barry White', who was known for his baritone voice, to 'Kylie Minogue'.

In addition to this, Carrie talks about having 'more fun' with her voice around people she is comfortable with. When I specifically asked her about how she though her voice changed between the two recorded conversations, she responded by saying that she thought she was using her voice 'more like a musical instrument' and noted that there was 'an element of performance', because she was trying to make Astrid laugh (Quote 5). She says that with Astrid she's 'less self-conscious' which in turn allows her to be 'more enthusiastic' and 'more foolish' in how she speaks.

Because of these comments that Carrie makes about her voice, I expected that she might use an increased pitch range in the conversation with Astrid. I return to this point in Section 12.4, where I compare Carried voice between the two recorded conversations.

12.2.3 'Phone voice'

Carrie uses terms like the 'phone voice', the 'learned voice', the 'all-the-people-that-are-gonna-misgender-me voice', the 'high voice' and 'the voice' throughout the interview,

which all appear to refer to a similar concept: A higher pitched voice learned in voice therapy that she uses to avoid misgendering, harassment and awkwardness. In the IPA, we saw that her decision to use this voice depended was shaped by external pressures (Theme B.1, Quote 16, and that she was growing increasingly resistant to feeling as if she should use it to make others more comfortable (e.g. 30).

She talks about the goals of voice therapy, which primarily involved increasing pitch, but also resonance and intonation. Pitch increases appear to have been the primary goal: 'We were using like an ipad with pitch detection to show me where I was and then I would try and move the range up'.

However, intonation and pitch range seemed to be play an important part too:

(58) There's significant differences between how stereotypically men and women speak, you know. There's much more variety there's much more life, whereas men tend to be more monotonous and I found that I'd been doing that but the I did try it and I was doing that thing of I'd be in the car and I'd be listening to phone-ins on the radio and I would try and copy the pitch and delivery of the women who were phoning in.

Carrie also produces the examples of her 'phone voice' during the interview, such as in quote Quote 29 includes an example of this voice. As shown in Figure 12.6, taken from Quote 29, an increase in pitch is a main feature of this voice, with her f0 in this section reaching 400 Hz (shown in green).

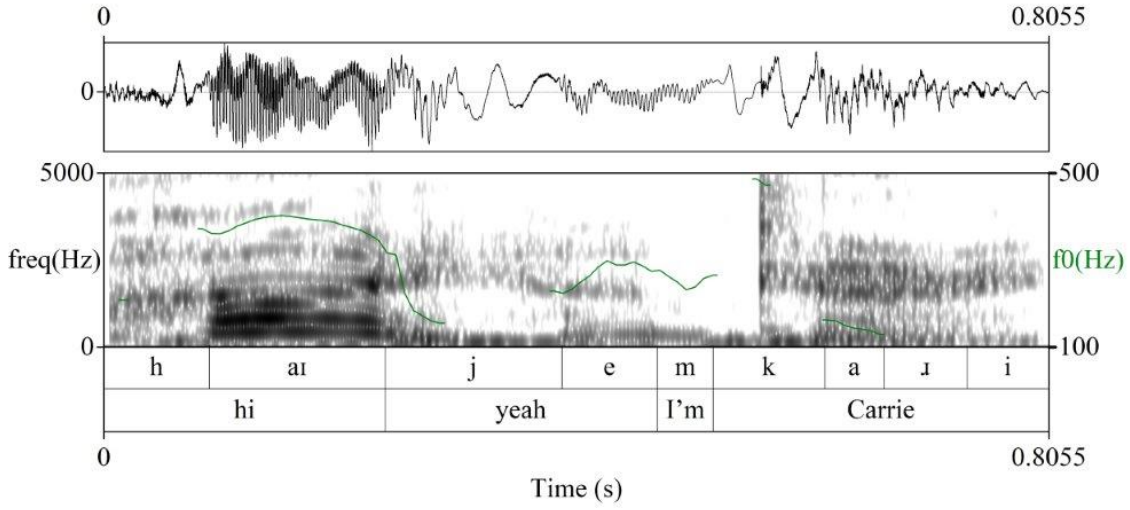


Figure 12.6: A wideband spectrogram showing Carrie's high pitch in her phone voice

The role of intonation and pitch range are also seen in her comparison of this voice to speaking like 'a children's TV presenter'. She also mentions resonance briefly, but does not give any more details on what this involved.

When I asked if 'phone voice' was the same as her 'radio voice', she described the difference primarily in terms of their purpose: 'My phone voice is me trying to stop the red mist from descending!'. The difference is that it is 'less playful', whereas her radio voice is 'more of a performance'.

In Figure 12.7, taken from a section where she is discussing the voice she was learning in speech therapy, we also see another component: Varied intonation and a high degree of pitch fluctuation. Carrie's f0 here goes from a low of 83 Hz to a maximum of 218 Hz, with the pitch contour shown in green showing rapid upwards and downwards movements.

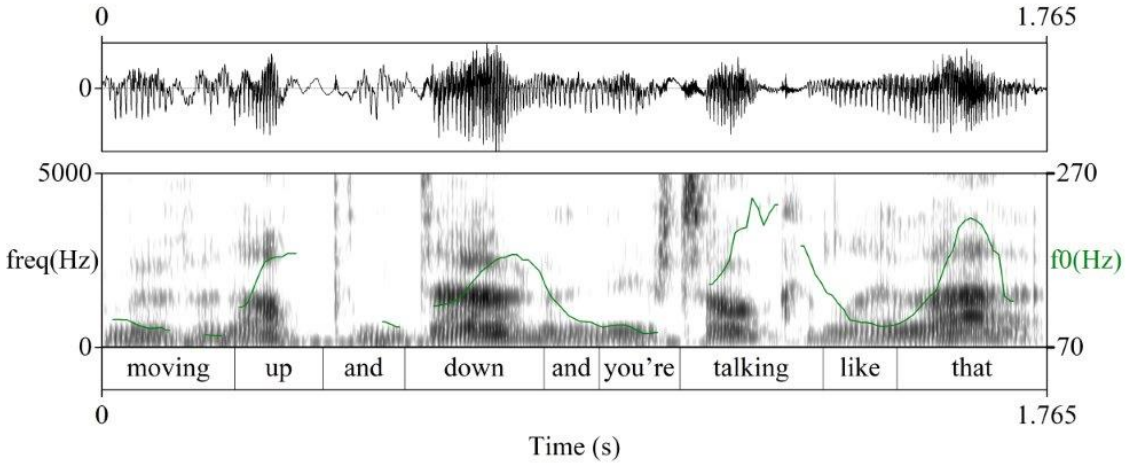


Figure 12.7: A wideband spectrogram showing how Carrie's f0 moves up and down as she imitates the voice she produced in voice therapy

12.2.4 The 'man voice' and lower pitch

Carrie variably refers to a version of her voice 'the deep man voice' and 'the man voice'. At times, this is her way of describing her voice generally: 'I still think I've got a great voice, but I just wish it wasn't a man voice'.

However, elsewhere, she uses the term to refer to particular instances of her voice where it has a lower pitch than usual, separating particular times where she has used a low pitch from her voice more generally. These shifts can be unintentional ('I keep catching myself going into the like [low pitch shift] "the deep man voice" and catching myself and pulling it back up') or intentional ('I've deliberately used the man voice while presenting female just to tell people to get the fuck out of my way', 'I have this real fear of the man voice being used').

She also discusses instances of using times where her voice is lower in pitch without using the term the 'man voice' in other instances: 'If I get a cold or I've been doing a lot of singing my voice goes lower for several days afterwards'.

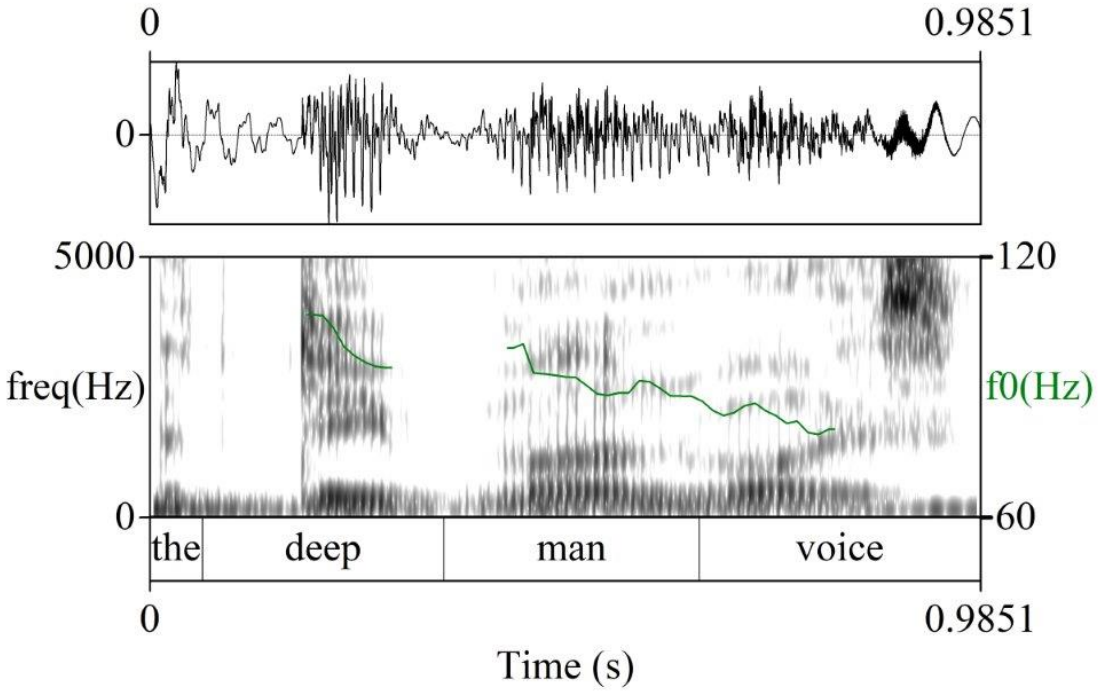


Figure 12.8: Carrie's production of 'the deep man voice'

Carrie produces an impression of what she calls 'the deep man voice', shown in Figure 12.8. Low pitch is a component of this, with a median pitch of 89 Hz compared to her overall median of 103 Hz in the conversations, but limited variation in pitch also appears to be an aspect.

12.2.5 Singing

Carrie compares her voice to the lead singers of several rock bands, making comparisons to Thom Yorke from Radiohead, Lemmy from Motörhead and Kurt Cobain from Nirvana. Thom Yorke is known for his wide vocal range and use of high pitch and falsetto, and his voice has been described as, 'high, keening sound, often trembling on the edge of falsetto' (Rolling Stone 2008). On the other hand, Lemmy's voice has been described as 'A voice like shrapnel and a bass tone to match' (Moyer 2015), and as 'ground up gravel in his throat' (Hartmann 2019).

She discusses how her singing voice has changed since coming out:

(59) Bands I was in to begin with, it was like, I would be sent back into the vocal booth to do it more rocky, because as far as they were concerned there was only one voice a male singer should have and it was like [harsh voice] 'ehnnnng', right- kind of thing. Um, so there's- I do a lot more in terms of harmony and a lot more in terms of pitch, like a lot of things we

> do now it's like, uh, there's like six or seven 'me's doing all the different bits and interacting.

In this example, Carrie uses a harsh voice to convey a part of her 'rocky' singing style. As shown in Figure 12.9, this kind of harsh voice is consistent with 'laryngeal constriction at high pitch' described by Esling et al. (2019). F0 reaches almost 500 Hz, though both examples seem to involve multiple periodicities and a high degree of noise, making it difficult to accurately estimate f0. She produces another example of this in the example shown in Figure 12.10.

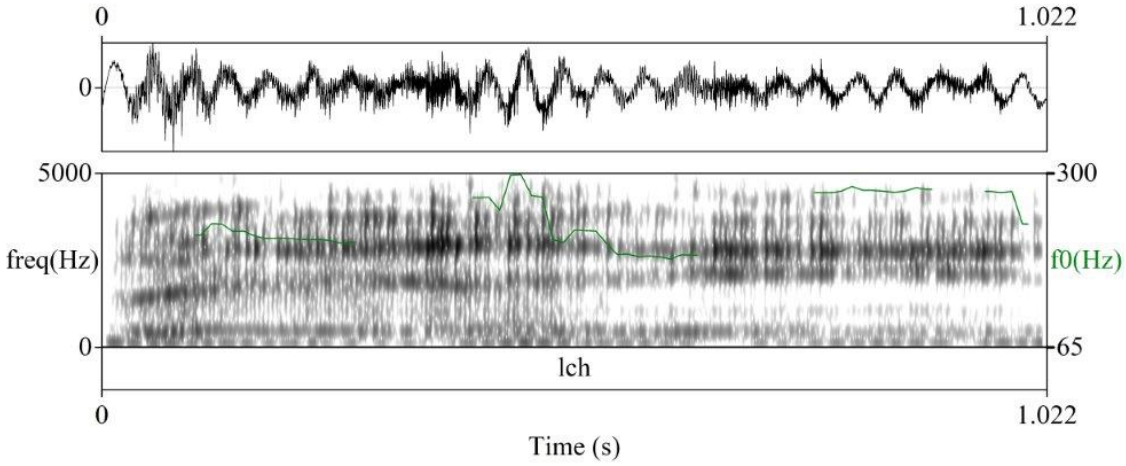


Figure 12.9: Carries use of a high-pitched form of harsh voice, laryngeal constriction at high pitch (annotated as lch), to convey her 'rocky' singing style. Max f0: 498 Hz, Min f0: 337 Hz, median f0: 396 Hz

12.2.6 The 'beast' voice

Elsewhere, Carrie uses harsh voice to convey the way that she perceives others to see her presentation and voice as incongruent. Carrie never gives a specific name to this voice, so I draw on the example in Quote 23 to term it the 'beast' voice. In Quote 23, she used this voice to create a contrast between her voice and that of her cisgender co-host on the radio, saying 'there's a *Beauty and the Beast* thing going on here!'.

This example is shown in Figure 12.11.

Two further examples come from the same story, told once in the interview and once in the conversation with Astrid, where she recounts how she befriended a group of strangers at an open mic night, and one of them said to her 'I didn't realise you were trans until you opened your mouth' and then did an impression of her, which she uses harsh voice to produce. As can be seen in Figure 12.12, these periods are characterised by a high degree of aperiodic noise. Though the noise interferes with any pitch tracking attempts, manually calculated f0 goes as low as 40 Hz in both samples.

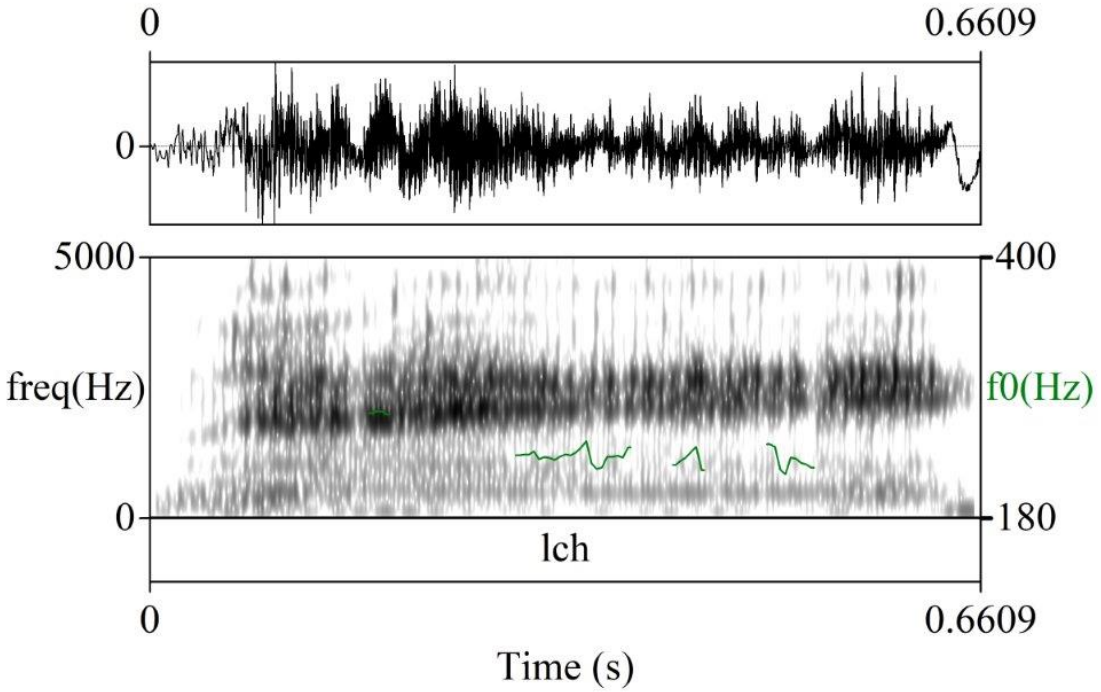


Figure 12.10: Carries use of a high-pitched form of harsh voice, laryngeal constriction at high pitch (annotated as lch), to imitate 'singing with no technique'

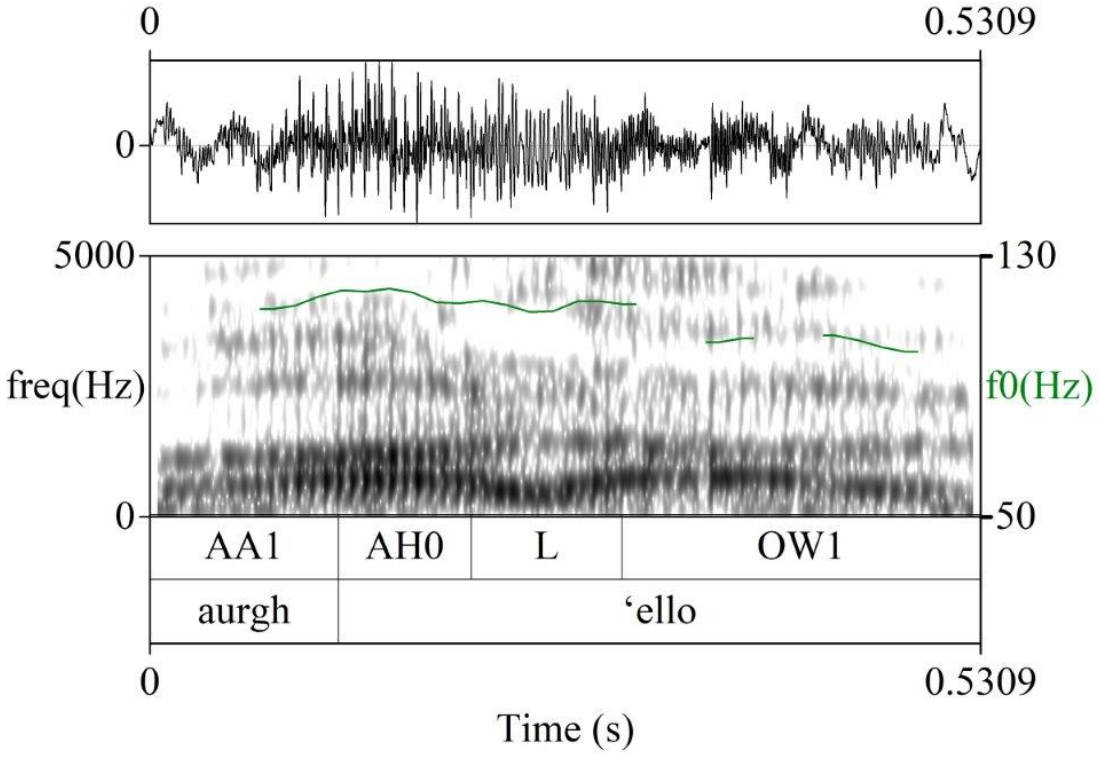
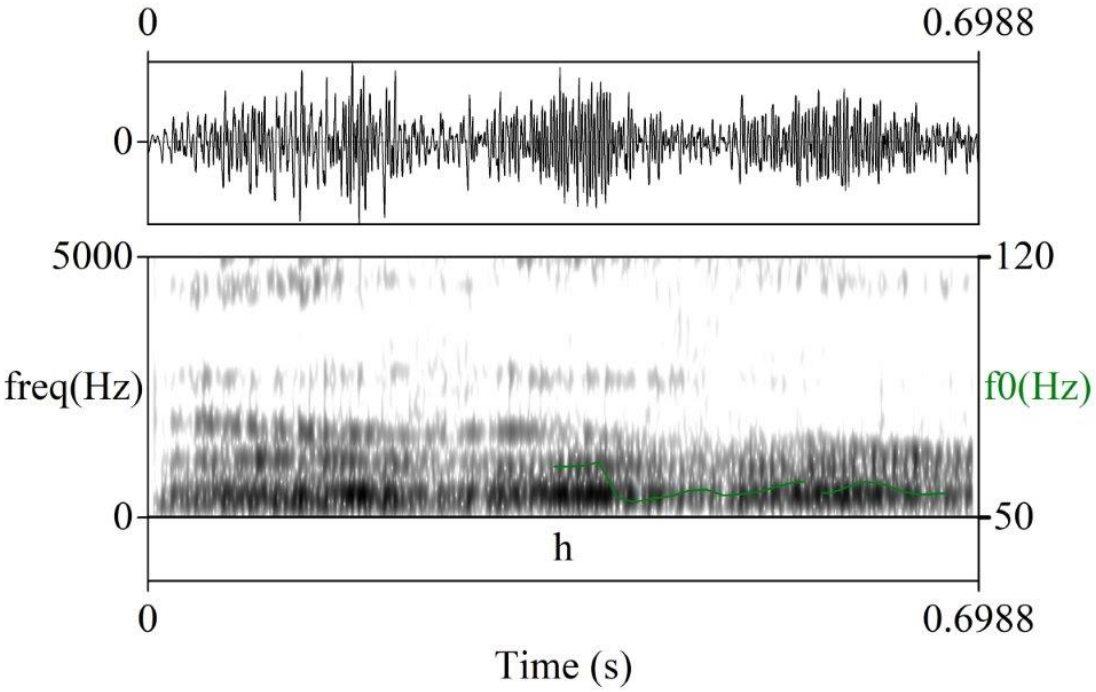
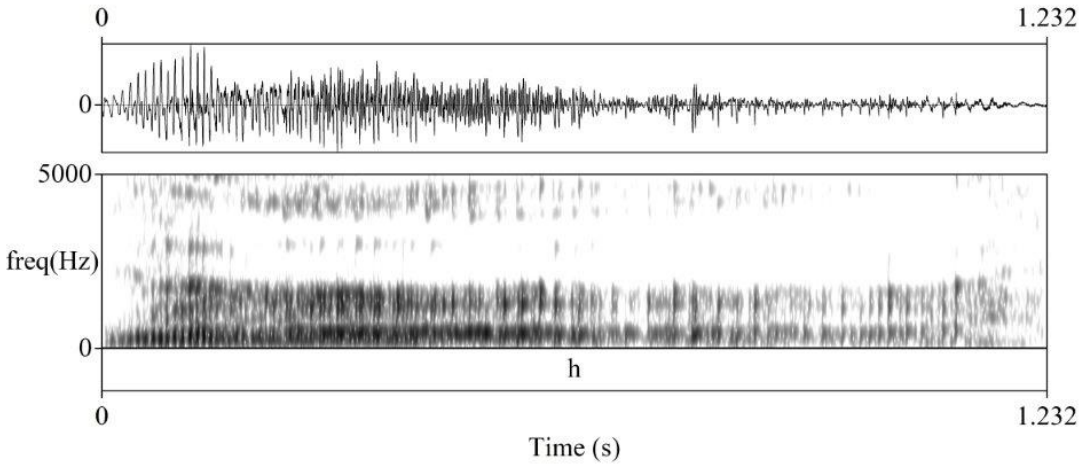


Figure 12.11: Carrie's production of harsh voice from the *Beauty and the Beast* example. [name] will say her thing and then I'll go [harsh voice] *'aurgh 'ello!'*

This type of harsh voice is thus distinct from the 'laryngeal constriction at high pitch' voice used to convey her signing style.



(a) From the conversation with Astrid.



(b) From the IPA interview

Figure 12.12: Carrie's productions of harsh voice from the open mic story

Carrie also does use 'beast voice' to convey how she believes audiences perceive her singing voice, however. In Figure 12.13, taken from quote 60, Carrie produces a low pitched harsh voice to frame how she believes her audience perceives her voice in contrast to her presentation when she is one stage, with f0 here dropping as low as 50 Hz.

(60) Which didn't fit with the presentation, 'cause I was dressed nice, y'know - I was dressed for prosecco and pizza night, I think, and- and this- this

[harsh voice] came out

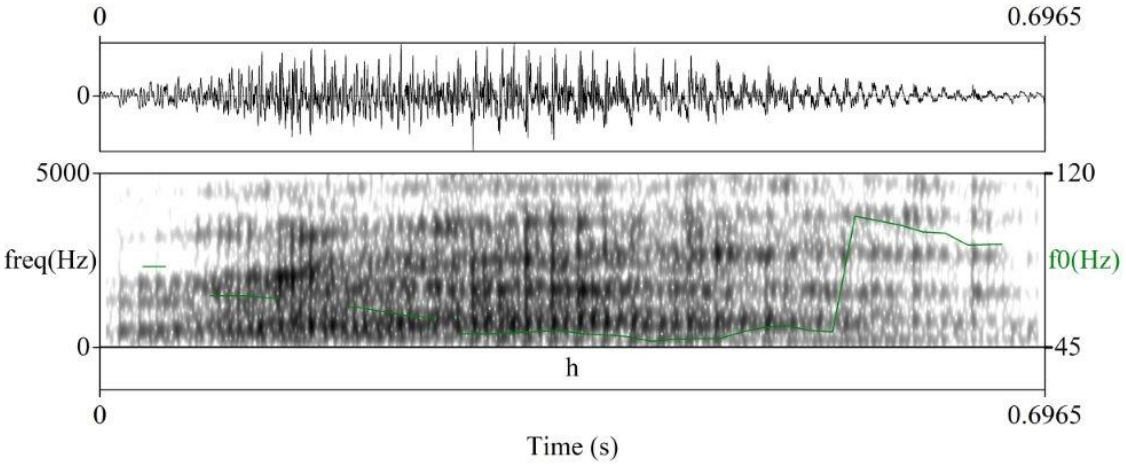


Figure 12.13: Carrie's production of harsh voice used to convey how she believes audiences to perceive her singing voice

Carrie often juxtaposes examples of the 'beast voice' next to descriptions of normative femininity. This can be seen in the original 'Beauty and the Beast' example, as well as in Quote 60. This happens again in the example shown in Figure 12.14 comes from Quote 8, where she talks about being 'out doing the high voice' but not being able to sustain this when coughing, and imitates this as shown in Figure 12.14.

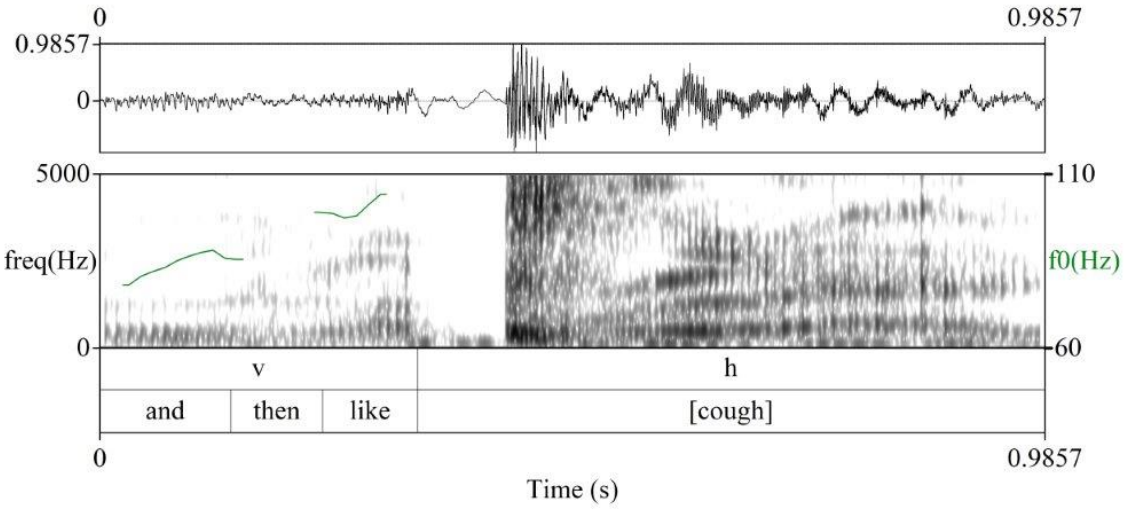


Figure 12.14: Carrie's production of harsh voice in the imitation of her cough

12.2.7 Not talking

Carrie talks about avoiding using her voice in situations that people may notice a disconnect between her voice and how she presents and where she feels in danger because of this. This is discussed in more detail in Section A in terms of the role that

Carrie's control over the situation and reactions of others plays in this. Carrie gives an example of this in a subway carriage, and public transport is a typical place that this occurs: 'On public transport, I just generally try not to talk'.

12.2.8 Contrast with linguistic use of non-modal phonation phonation types

Although we have already seen cases where Carrie uses laryngeal constriction at high voice to convey her singing voice, and low pitch harsh voice to convey her 'beast' voice, there is one final context where Carrie uses harsh voice quality. The final context in which Carrie uses harsh voice is to add emphasis. Shown in 12.15 and 12.16. These instances are clearly distinct in that they occur very briefly on lexical items rather than in isolation or in constructed dialogue, occur on intensifiers, and are also very whispery.

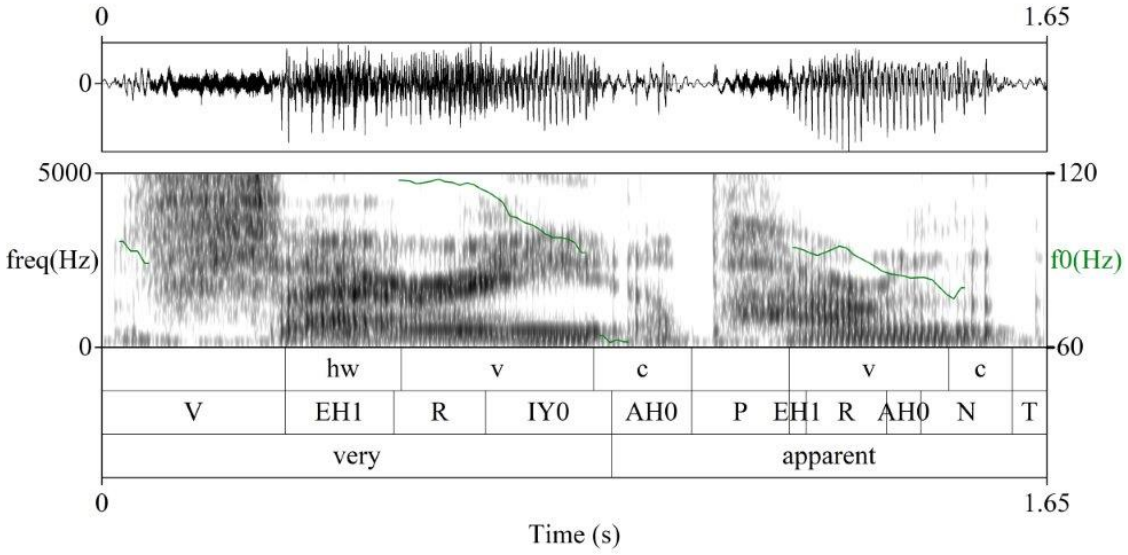


Figure 12.15: Harsh whispery voice used for emphasis to produce in 'very'

Carrie also uses whisper for parentheticals (Figure 12.17a) and when backchannelling (Figure 12.17b. Furthermore, the example in Figure 12.17c shows a more linguistically-conditioned use of whisper, where Carrie uses whisper in an initial vowel onset, which appears to be an alternative to a glottal onset.

Carrie's use of creaky voice appears to be overwhelmingly conditioned by linguistic factors. For example, Figure 12.18 gives examples of creak occurring briefly as a glottal stop and glottal onset.

Carrie rarely uses extended creak over multiple consecutive syllables, but does do so in the following extract, shown in Figure 12.19:

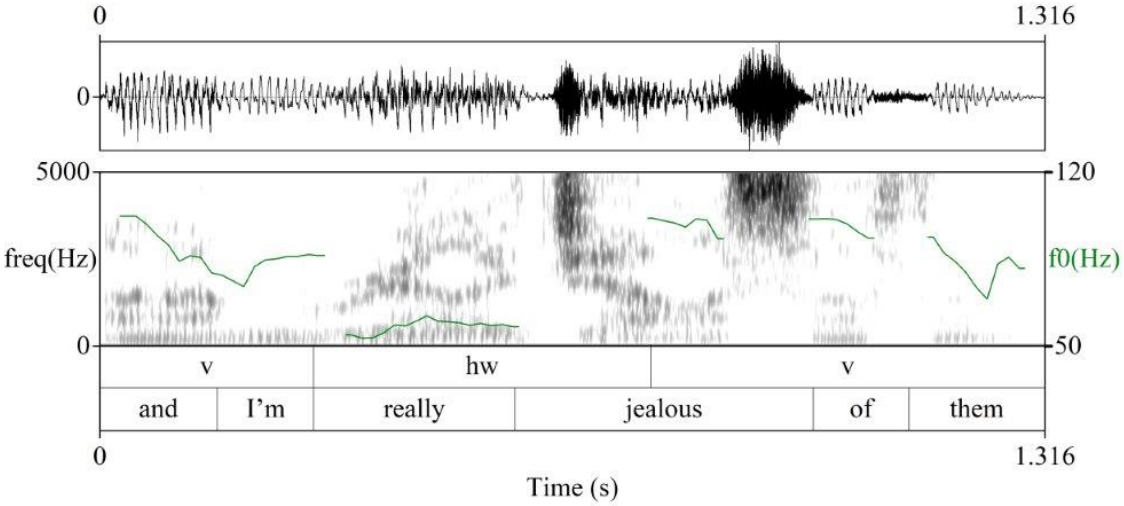
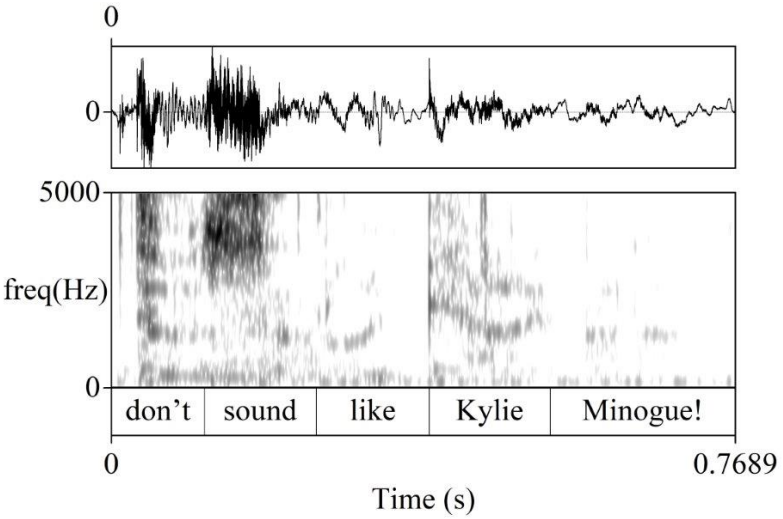


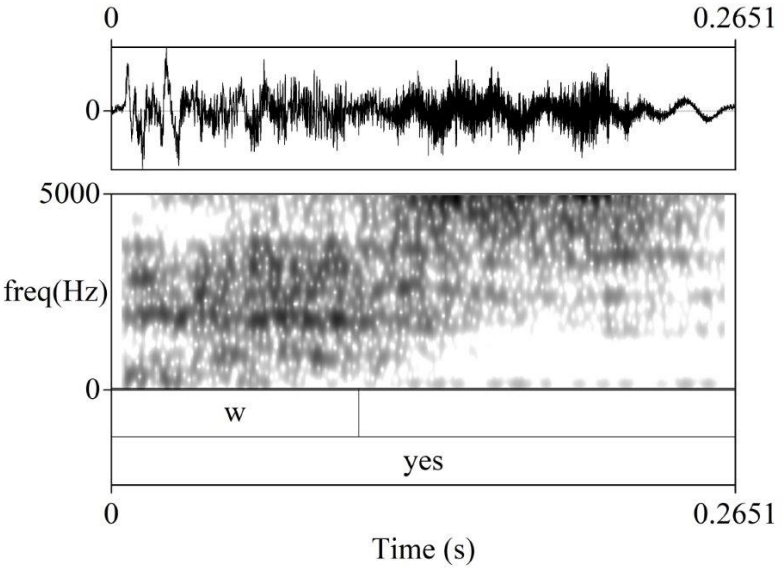
Figure 12.16: Harsh whispery voice used for emphasis in the word 'really'

(61) There's this idea that if you come out, particularly in later life, you're gonna be run out of town, you know with yokels with flaming torches and stuff and it was- always part of me was expecting trouble, um and it's taken quite a while before I've understood that what's in your head and what's in the internet are not necessarily what's happening in real life

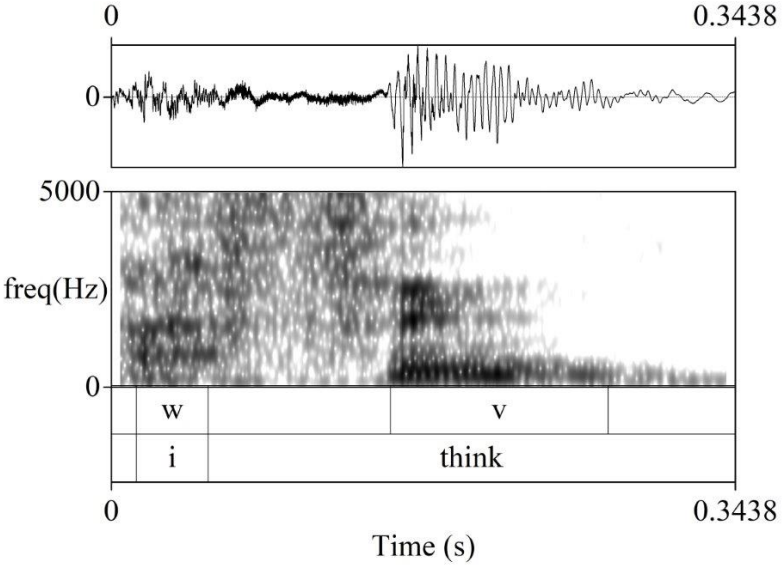
Here, Carrie is reflecting on her fears of other people reacting negatively to her being trans. Her use of creak here suggests that she may be using creak to convey negative affect.



(a) Carrie's use of whisper in a parenthetical

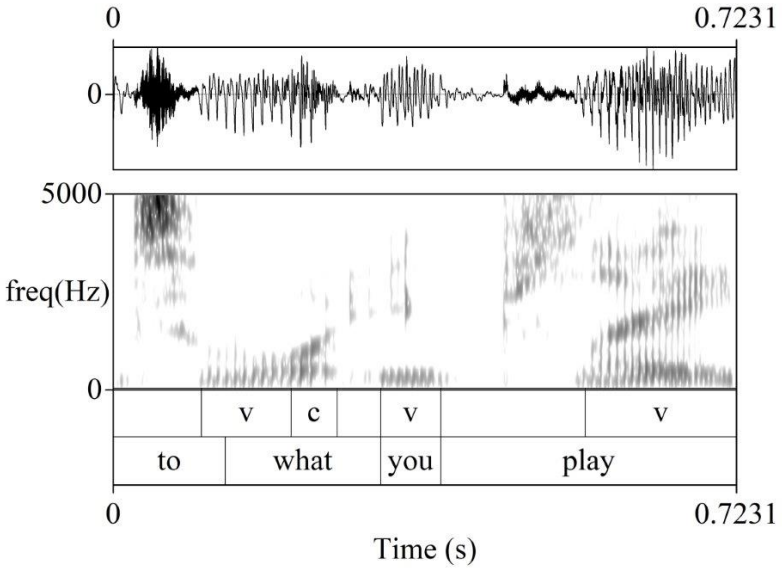


(b) Carrie's use of whisper to backchannel

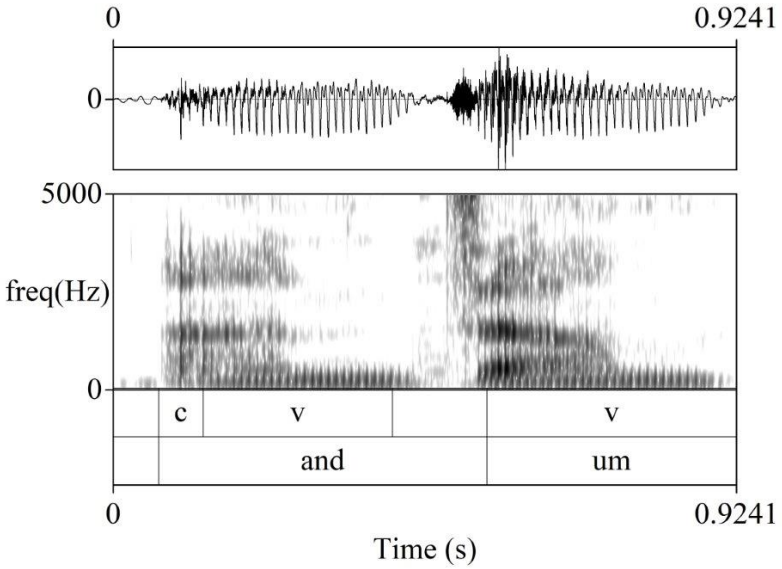


(c) Carrie's use of whisper in an initial vowel onset

Figure 12.17: Carrie's use of whisper in three contexts



(a) Carrie's use of creak to produce a glottal stop



(b) Carrie's use of creak in a glottal onset

Figure 12.18: Carrie's use of creak conditioned by linguistic factors

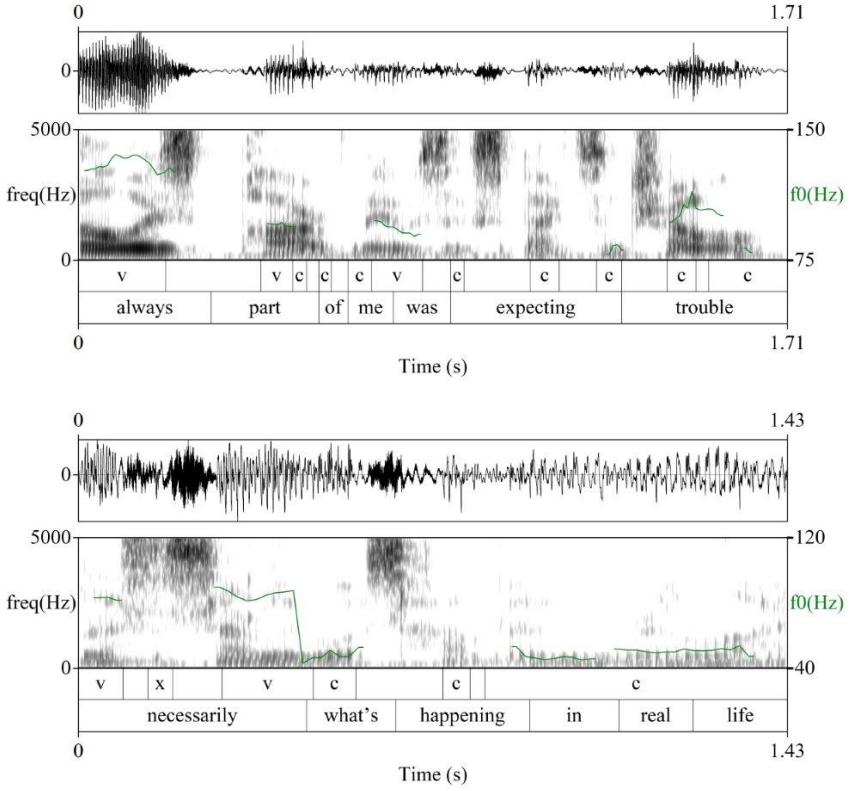


Figure 12.19: An example of Carrie using creak over multiple syllables to convey negative affect

12.3 Quantitative analysis of Carrie's voice quality

In this section, I present the results of the quantitative analysis of Carrie's voice in the two recorded conversations, one speaking to her friend Astrid, and one speaking to an unknown interlocutor, Jane. When discussing her voice in the IPA interview, she told me that she aimed for 'a more breathy voice than God intended' but said that 'I'm very aware it's not remotely breathy today'. Considering the links between breathy voice and femininity that circulate in media (Jeong 2017) and the fact that breathy voice often occurs as a target in feminising voice therapy (Davies, Papp & Antoni 2015), I postulate that this 'more breathy' style may related to some aspects of the high-pitch 'learned voice' just described in Section 12.2, which Carrie uses in situations where she aims to avoid being misgendered. As Carrie was recorded speaking to Jane, an unknown interlocutor, I expected that she might incorporate aspects of this into her voice in the conversation with Jane

As seen in Section 12.2.2, Carrie also described how she used her voice in the conversation with Astrid, which involved being able to incorporate a wider range of gendered styles and being able to incorporate an element of performance into her voice, to use it 'like a musical instrument' and be 'more foolish'. I therefore expected that Carrie would display a wider pitch range in her conversation with Astrid.

12.3.1 Analysis of categorical phonation types

To recap the methods used, rather than coding according to PPA, I instead coded for cases where Carrie's voice quality categorically deviated from a 'baseline' quality. This 'baseline' included modal, breathy, whispery and tense voice quality. I did not follow a 100 ms cutoff point. After excluding coding errors, a total of 4336 stretches were coded in Carrie's speech.

Phonation type	n	% phonation type
Baseline voicing	3232	74.5%
Creak	1066	24.6%
Whisper	35	0.8%
Harsh	3	0.1%
Total	4336	100.0%

In 74.5% of voiced stretches, Carrie's voice covered a range of modal, breathy and whispery voice, which I term here as 'baseline voicing'. Carrie made use of creak, whisper, and forms of harsh voice. Of these, creak was the most common phonation type used, accounting for 24.6% of voiced stretches. Due to the difference in methods, this is unfortunately not comparable to the proportion of creak used in the corpus study previously: most of this creak occurs for very short periods of time, and would

	Min	Q1	Median	Mean	Q3	Max
f0	41.48	91.57	102.73	111.41	121.62	409.40
H1*–H2*	-21.90	1.30	3.42	3.36	5.49	15.80
H2*–H4*	-11.54	2.57	5.19	5.36	8.16	20.35
H4*–2kHz*	-19.86	1.95	5.40	5.37	8.77	47.74
CPP	13.22	16.64	17.66	17.78	18.80	25.20

Table 12.3: Descriptive statistics of Carrie's f0, H1*–H2*, H2*–H4*, H4*–2kHz and CPP overall.

have been excluded in the PPA analysis presented in Chapter 6. Whisper and harsh voice were used comparatively infrequently, both in less than 1% of cases, and so these are not included in the present quantitative analysis. Instead, I considered these in the integrated qualitative-acoustic analysis presented in Section 12.2.

I carried out a mixed-effect logistic regression model to consider the effect of interlocutor on Carrie's use of creak. However, this analysis produced no significant effect for interlocutor, and the results can therefore be found in Appendix C.6.

I then considered how Carrie's voice varied between the two interviews in terms of acoustic measures.

12.4 Descriptive statistics

Descriptive statistics for H1*–H2*, CPP, and f0 in stretches with baseline voicing provided in Table 12.3 to aid in interpretation of results.

Figure 12.20 shows how the range of Carrie's f0 varies between the two conversations. Though the mean and median f0 of Carrie's voice is very similar in each conversation (Astrid median = 103 Hz, mean = 114 Hz;Jane median = 103 Hz, mean = 109 Hz), she has a higher maximum and lower minimum f0 with Astrid than with Jane (Astrid min = 41 Hz, max = 409 Hz; Jane min = 44Hz, max = 380 Hz) as well as a wider interquartile range (IQR Astrid = 39 Hz, IQR Jane = 24 Hz).

I carried out a series of mixed-effect linear regression models to consider the effect of interlocutor on H1*–H2*, H2*–H4*, H4*–2kHz* and CPP. However, H2*–H4* and H4*-2kHz* produced no significant effect of Interlocutor. These can be found in Appendix C.5. Here, I present only the results of the linear mixed-effects models predicting H1*–H2* and CPP as a function of linguistic factors and interlocutor.

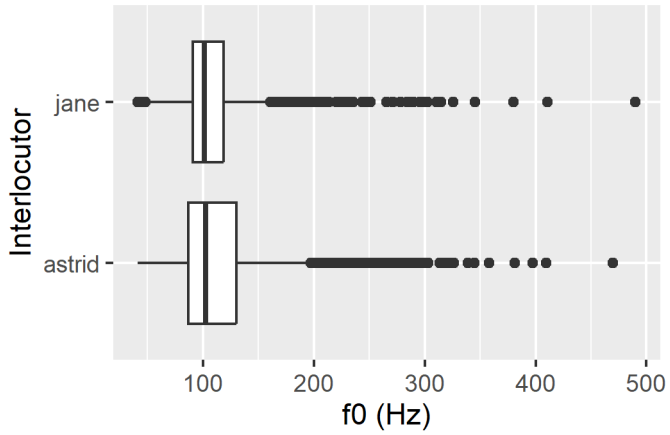


Figure 12.20: A boxplot show how Carrie's f0 range varied between the conversation with Jane and the conversation with Astrid

12.4.1 H1*–H2*

To recap, H1*–H2* is often used as a measure of glottal constriction, where higher values relate to laxer and progressively breathier phonation, while lower values relate to more constricted phonation. In the Part II, I found that whispery voice, modal voice, tense voice and tense whispery voice occupied the middle range of H1*–H2*, while increased H1*–H2* related to breathy voice and lower H1*–H2* related to tense voice.

Here, I analyse H1*–H2* as a function of Duration, Aspiration, Phrase Position and Interlocutor in a mixed-effects linear regression model.

There was no evidence of problematic collinearity in the model with all GVIFs $<$ 1.5.

Groups	Name	Variance	Std.Dev.
words	(Intercept)	0.079615	0.28216
words.1	Duration	0.003087	0.05556
Residual		0.736709	0.85832

Table 12.4: Number of obs: 3206, groups: words, 1211

Table 12.5 shows the results of the mixed-effects linear regression predicting H1*–H2* as a function of Duration, Aspiration, Phrase Position, Vowel, and Interlocutor. The effect of random effects is given in Table 12.4

Predicted H1*–H2* decreased with longer stretch duration, indicating that longer stretches tend to be produced with tenser phonation.

Aspiration from preceding /p,t,k/ and aspiration from preceding /h/ both increased predicted H1*–H2*, relative to stretches with No Aspiration from preceding segments.

Table 12.5: Results of the mixed-effect model considering H1*–H2* in Carrie's baseline voice quality as a function of interlocutor and linguistic factors

Independent variable	Level/unit	Coefficient (SE)
Intercept		-0.401^{***} (0.031)
		t = −12.861
Duration	Log transformed and scaled	-0.072^{***} (0.020)
		t = −3.683
Aspiration	Aspiration from preceding /p,t,k/	0.152^{***} (0.040)
(Ref = No aspiration)		t = 3.749
	Aspiration from preceding /h/	0.326^{***} (0.095)
		t = 3.426
Phrase position	Final	0.091^{*} (0.040)
(Ref = Not final)		t = 2.287
Vowel	Initial	0.593^{***} (0.078)
(*Ref = Non-initial vowel*)		t = 7.555
	Both	0.375^{*} (0.146)
		t = 2.563
	None	−0.018 (0.151)
		t = −0.116
Interlocutor	Jane	0.512^{***} (0.032)
(*Ref = Astrid*)		t = 15.764
Observations	3,206	
Log Likelihood	−4,203.589	
Note:	*p<0.05; **p<0.01; ***p<0.001	

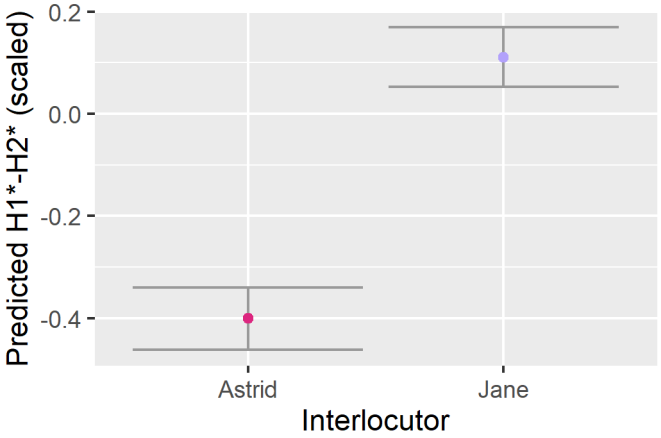


Figure 12.21: The effect of interlocutor on predicted H1*–H2* in Carrie's voice. Error bars represent 95% CIs.

Compared to occurring in non-final position, occurring in Phrase Final position increased predicted H1*–H2*.

Relative to stretches that contained of Non-Initial Vowels, stretches that contained Initial Vowels and Both initial and non-initial vowel increased predicted H1*–H2*.

As shown in Figure 12.21, H1*–H2* is higher in the conversation with Jane than in the conversation with Astrid, indicating laxer phonation when speaking to Jane.

Table 12.6: Pseudo R^2 measures for baseline and full model for H1*–H2*. Baseline model include only a random effect for Words

Model	R2m	R2c
Full	0.1	0.19
Baseline	0	0.13

12.4.2 CPP

To recap, CPP is a measure of aperiodic and harmonic noise in the signal, where higher values represent increased periodic noise, and lower values represent increased aperiodic noise. In Part II, I found that breathy, whispery and tense whispery voice showed low CPP, while modal and tense voice showed high CPP.

Here, I analyse CPP as a function of Duration, Speech Rate, Aspiratoin, Vowel, Contains Nasal and Interlocutor.

As Duration increased, predicted CPP increased.

As Speech Rate increased, predicted CPP decreased.

Following /h/ increased predicted CPP.

Table 12.7: Results of the model considering CPP in Carrie's baseline voice quality as a function of interlocutor and linguistic factors

Independent variable	Level/unit	Coefficient (SE)
Intercept	0.011 (0.036)	
		t = 0.313
Duration	Log transformed and scaled	0.243*** (0.023)
		t = 10.419
Speech rate	Log transformed and scaled	−0.062** (0.021)
		t = −2.952
Aspiration	Aspiration from preceding /p,t,k/	−0.058 (0.043)
(Ref = No aspiration)		t = −1.359
	Aspiration from preceding /h/	0.283** (0.101)
		t = 2.807
Vowel	Initial	−0.321*** (0.083)
(*Ref = Non-initial vowel*)		t = −3.884
	Both	−0.264 (0.155)
		t = −1.698
	None	−0.108 (0.159)
		t = −0.678
Contains nasal	Nasal	−0.147** (0.046)
(*Ref = No nasal*)		t = −3.178
Interlocutor	Jane	0.125*** (0.034)
(*Ref = Astrid*)		t = 3.621
Observations	3,206	
Log Likelihood	−4,397.667	
Note:	*p<0.05; **p<0.01; ***p<0.001	

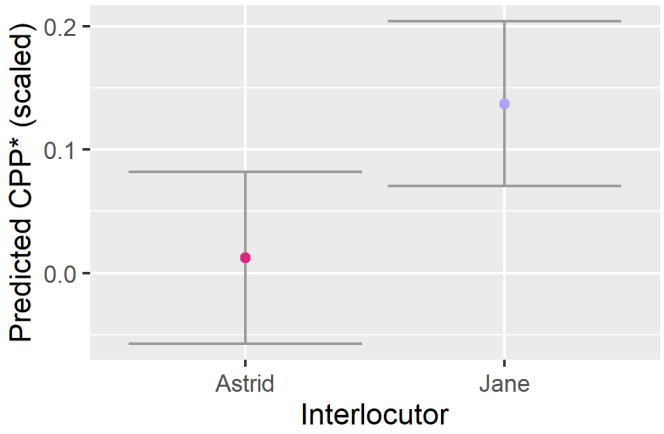


Figure 12.22: Effect of Interlocutor on predicted CPP

Model	R2m	R2c
Full	0.07	0.16
Baseline	0	0.1

Table 12.8: Pseudo R^2 measures for baseline and full model for CPP. Baseline model include only a random effect for Words

Compared to stretches containing non-initial vowels, stretches that contained Initial vowels decreased CPP. Containing a Nasal decreased predicted CPP.

As shown in Figure 12.22 predicted CPP differed significantly between the two Interlocutors: In the conversation with Jane, predicted CPP was higher than in the conversation with Astrid, indicating more modal-like phonation in the conversation with Jane.

Chapter 13

Discussion

I conducted an Interpretative Phenomenological Analysis of a semi-structured interview with Carrie in an attempt to understand how she makes sense of her experiences with her voice. I identified three overarching themes: (A)The importance of control; (B)Feelings around voice training and codeswitching: A tension between self-perception, social pressure and fear; (C) Self-acceptance of voice and self as the way forward. I also found that she made reference to a number of distinct 'voices' that she used in her life (e.g. 'work voice', 'man voice') characterised by shifts in pitch and voice quality. I then considered how her voice changed between the two conversations, and found that H1*–H2* and CPP were significantly higher with Jane, an unknown interlocutor, than with Astrid, a close friend.

Here, I bring together the quantitative and qualitative analyses and discuss them in the context of theory from sociolinguistics and transgender studies. I begin by discussing the acoustic nature of Carrie's voice quality in the context of Scottish voice quality, relating the findings back to those presented in Section 8.3. I then discuss IPA findings, bringing the qualitative and quantitative acoustic analyses where relevant. In accordance with IPA, which allows themes to arise in the analysis that were not anticipated by the researcher, I also bring in texts were not explored the background to this research.

13.1 Acoustic nature of Carrie's voice

After separating cases of harsh voice, whisper and creak to facilitate interpretation of acoustic analysis, I considered how Carrie's baseline voice quality manifested acoustically and compared how her voice varied between two conversations.

Measurements for H1*–H2* in Carrie's voice centred around a mean and median of 3dB, with the middle 50% of the data ranging from 1.30dB (Q1) to 5.49dB (Q3), though her full range for H1*–H2* extended far beyond this (–21dB to 16dB). If we compare this to the findings for how PPA-coded voice quality manifests acoustically, presented in Section 8.3, we see that Carrie's voice tends to be close to centre around the mean of 4dB for H1*–H2* for Scottish voices more generally.

In terms of H2*–H4*, we see that Carrie's voice shows a mean and median of 5dB, with the middle 50% of her voice ranging from 3dB to 8dB. If we compare this to how voice quality manifests acoustically in Scottish voices more generally, we see that Carrie's voice tends towards modal-breathy values for H2*–H4*; the mean for Scottish voice quality in the smaller corpus was 3dB, and an increase to this significantly increased the chance for a voice being categorised as breathy.

In terms of H4*–2kHz*, we can see that Carrie's voice centres around a mean and

median of 5dB, with the middle 50% of the data ranging from 2dB to 9dB. This is close to the mean of smaller SCOSYA corpus of 5dB, suggesting a tense whispery or whispery quality.

In terms of CPP, Carrie shows a mean of 18dB, with the middle 50% of her voice ranging from 17dB to 19dB. This is considerably lower than 21dB mean of the data for the smaller SCOSYA corpus presented in Section 8.3, especially considering that Carrie's data was recorded in a speech studio rather than in a home environment.

Taken together, I suggest that these values suggest that the combination of high H2*–H4* and low CPP in comparison to the values shown in SCOSYA suggest that Carrie's voice generally tends towards a breathy quality. This is also consistent with my auditory impression of her voice. However, the proximity of her voice to the mean of Scottish voice quality in terms of H1*–H2* and H4*-2kHz* suggests that her voice may also involve a whispery component.

Carrie's voice also shifted between the two conversations. Compared to the conversation with Jane, Carrie showed increased pitch range when talking to Astrid. Meanwhile, when talking to Jane, she showed higher H1*–H2* and higher CPP than when talking to Astrid. I therefore suggest that Carrie's voice quality ranges from breathy to whispery in spontaneous speech, and specifically that it becomes laxer and more modal-like in conversation with Jane. This shifting is discussed in more detail in the next section.

13.2 Control, agency and context

The first overarching theme highlighted the importance of control in Carrie's use of and understanding of her voice, involving both control over her own voice and the context in which she used it. Carrie's expert control over her voice and the degree of control that she has over a situation are key to explaining how she uses her voice in different situations.

Zimman (2021: 85) notes the role of agency as a potential avenue for future research on trans speech, noting that how speakers construct their own agency may play a role in their sociolinguistic behaviour. The results of the IPA highlight this here: While she demonstrates a high degree of vocal expertise, Carrie considers there to be certain immutable aspects of her voice ('you can't change your cough' (Quote 8, 'I can't make my vocal chords magically change' (Quote 43). For Carrie, the fear that a lack of control brings affects how she uses her voice.

Carrie's decision to avoid speaking in certain contexts, such as on public transport,

may be a way to regain some control over a situation. Babel (2016) explores how silence can be a form of control in data from an ethnographic study in Santa Cruz valleys of Bolivia where she considers women's participation in public meetings, which carry the expectation to speak in *oratoria*, a particular genre of speech used in formal situations. Babel (2016) explores how women remain silent in meetings because of an awareness that they are not seen as capable of public speaking and instead participate in community decision making through off-the-record remarks, gossip, and whispered commentary of meetings. Babel (2016) argues that silence is a form of social action that women use to position the way that they are perceived. Carrie's use of silence in high-stakes situations appears to serve a similar purpose: By refusing to speak, she prevents her voice from being 'the giveaway', and avoids the consequences that come along with speaking with a voice that is perceived as incongruent with her presentation.

Many shifts that Carrie describes, between 'work voice', 'friends voice' and 'phone voice', can be interpreted within the framework of audience design (Bell 1984), where intraspeaker stylistic variation can be understood as a speaker's response to their addressee. In this framework, a style shift can be classed as either a responsive shift, where a speaker shifts in response to a change in audience (or another factor, such as topic or setting), or an initiative shift, in order to redefine an existing situation. Carrie describes shifting depending on her interlocutor, reporting distinct styles that she uses in singing, on the phone, and to friends. These shifts are not neat examples of either responsive or initiative shifts. Here, I explore how Carrie shifts her voice to redefine and assert control over situations, but does so in response to the expectations placed upon her by her audiences.

Performance is a key aspect of several of Carrie's voices. She describes raising her pitch and trying to speak more softly in her 'work voice', which she uses on the radio and in recording her audiobook, with the aim of creating intimacy with the listener. This is consistent with Bell's (1984: 192) description of initiative style design in media language, which Bell notes 'creates the relationship between communicator and audience, rather than responding to an existing relationship', emphasising that this is especially true for radio due to its lack of visual information. As Bell (1984: 170) states, when the size of audience increases, the addressee's influence on a speaker increases too. For Carrie, this manifests in a greater pressure to conform to the expected norms of female speech, namely, increased pitch ('There's an awful lot of bigots who are listening and I don't really want to give them the opportunity to go 'oahh'). Carrie also hears her own voiced played back to her in the studio, transforming her from speaker to addressee, asserting influence as a listener back onto her voice as a speaker ('I want my voice to be the most 'me' it can be [...] because [I'm] hearing it so loud').

Gibson & Bell (2012) explore how shifts in popular music singing can be interpreted as referee design, an initiative shift towards the style of reference group that a speaker

wishes to identify with. Carrie sent me a sample of her singing voice, which she described as 'an illustration of my extremely un-feminine voice when I'm singing', noting that the particular sample was 'probably more aggressive than usual'. On the one hand, aspects of this voice are initiative shifts towards a reference style, with the 'rocky' voice quality that Carrie describes involve 'laryngeal constriction at high pitch' as a convention of the genre. The extract that she sent me, not analysed in full here, also contain this in terms of segmental features, with her vowels shifting towards a General American reference style. However, this also needs to be interpreted with reference to the fact that Carrie takes joy in using this voice on stage because of she is diverging from listeners' expectations of her. Her control over both her voice and the situation when she performs on stage allow her to produce this initiative shift.

Carrie also sees her 'friends voice' as having 'an element of performance'. Consistent with predictions, this manifested in her using an increased f0 range in her conversation with Astrid. Among friends, Carrie feels less pressure to conform to gendered expectations in terms of pitch. Bell (1984: 169) describes how the relationship between speaker and addressee can have a role in stylistic variation. He suggests that among strangers, status is more important, while among close relationships, status is less important. The way that Carrie describes using her voice around friends can be interpreted within this framework: Just as there is less pressure to conform to ideals of standard language in close relationships, around friends there is less pressure for Carrie to conform to ideas of what women 'should' sound like.

Carrie's use of her 'phone voice', which involves an increase in pitch, can also be seen as an initiative shift to redefine an existing situation. Again, this shift is responding to the expectations of the listener. On the phone, Carrie has little agency over the situation, as seen in the example she gives of being repeatedly asked to put Carrie on the phone when trying to change something on her car insurance policy (Quote 45). By raising her pitch on the phone to reduce what she describes as 'intense awkwardness', she attempts to assert control over the situation, within the confines of the addressee's expectation that a higher pitched voice will accompany a woman's name.

Carrie does not appear to strictly use 'phone voice' in either recorded conversation, with her productions of it in the interview being higher in pitch than her average range of her conversational speech. However, her shifts in voice quality between the two interviews could potentially be seen as aspects of this. She noted trying to 'have a more breathy voice than God intended', revealing awareness of a link between female speakers and breathy voice. Carrie showed significantly higher H1*–H2* and CPP in the conversation with Jane, suggesting use of a lax, modal-like quality. Given the lower H1*–H2*, one interpretation of this is that she may use a breathier voice quality upon a first meeting with an unknown interlocutor, exploiting the connection between breathier qualities and femininity. This shift does not draw on the tenser quality used

by female speakers in Scottish accents found in corpus study presented in Part II. Instead, Carrie's use of a laxer quality might be connected to her understanding of a link between female speakers and a breathier voice quality, demonstrated through her assessment that she aims to be more breathy 'than God intended', presumably linking a less breathy quality to the effects of testosterone on laryngeal physiology.

However, an alternative explanation could be that when talking to an unknown interlocutor, Carrie styleshifts towards what she described as her 'radio voice'. Figure 12.4b shows an instance of Carrie's production of her 'work voice', which shows greater H1–H2 than in a nearby equivalent vowel. It is also notable that Carrie's voice also increases in CPP in the conversation with Jane, a trend which would be more consistent with a laxer, modal-like quality than a straightforwardly breathier one - or indeed a more whispery quality. Together, this may suggest that Carrie approximates her radio voice in conversation with Jane as an initiative shift, in an attempt to gain some control over the situation through vocal control.

13.3 The role of self-perception, social pressure and fear in Carrie's use of her voice

The second overarching theme explored tensions between different factors that motivating Carrie's vocal behaviour. The impact of wider societal attitudes towards trans people as a driver of linguistic behaviour has been identified in a number of previous studies. Hazenberg (2016) suggests that as trans speakers risk emotional and physical repercussions if their gender is called into question, they avoid using heavily gendered variants. Similarly, Podesva & Hofwegen (2015) suggest that the conservative sociopolitical context in rural California is relevant to trans speakers' linguistic behaviour. Carrie's own account of how she perceives and uses her voice demonstrates this similar concept, and here I draw on Carrie's account to discuss how wider sociopolitical attitudes towards trans people are essential for understanding trans speakers' linguistic behaviour with reference to her detailed accounts of her experiences.

The way that Carrie refers to her own voice and makes use of harsh 'beast' voice to convey how she understands others to perceive her own voice is best understood within the context of pervasive framing of trans women as both male and monstrous in discourse and media representations. On the one hand, Carrie refers to her voice as 'a man voice' and refers to particular instances of using a lower-pitched voice as 'the deep man voice', revealing an understanding of indexical links between low pitch and male speakers. Though it is likely common for speakers to be aware of an association between male voices and low pitch, for Carrie her use of the term 'man voice' appears to be connected to how gender is attributed to her voice by others, heightening her

awareness of this connection and allowing her to exploit it in her interview when she shifts her pitch downwards to produce 'the deep man voice'. Terming her voice a 'man voice' also allows her to distance herself from this label and this association, something that is also visible in her use of passive voice to describe instances of 'the man voice being used'.

Carrie's own view of her voice encapsulates wider discourse that frames trans women as being monstrous. Stryker (1994) discusses the connection between transsexual [1] women and monsters in a response to Daly (1978) and Raymond (1979). She notes that both Daly and Raymond link Frankenstein's monster and transsexual women's bodies. In a chapter entitled 'Boundary violation and the Frankenstein phenomenon', for example, Daly characterises transsexual women as 'dead matter molded into 'life-like' imitations of women, labelled 'The Real Thing" who engage in a 'necrophilic invasion' of womanhood. In response, Stryker (1994: 238) describes feeling an affinity to Frankenstein's monster:

> Like the monster, I am too often perceived as less than fully human due to the means of my embodiment: like the monster as well, my exclusion from human community fuels as deep and abiding rage in me that I, like the monster, direct against the conditions in which I must struggle to exist

Figure 13.1: Angry mob wielding flaming torches from Frankenstein (*Frankenstein* 1931)

In a parallel to Stryker's own affinity with Frankenstein's monster, Carrie's fears of how others will react to her voice appear are reminiscent of the scene in *Frankenstein* (1931), pictured in Figure 13.1 where the villagers with flaming torches hunt down Frankenstein's monster: 'My voice is gonna be the giveaway and that's when the flaming torches come out and they're gonna run me out of town' (Quote 7). Her perception of her voice as monstrous is also seen in her comparison of her voice to that of her cisgender radio co-host, where she says that there's 'a beauty and the beast thing

[1] I follow Stryker's usage of the term 'transsexual' in this paragraph

going on', using harsh voice to voice her understanding of how her voice is perceived in comparison to her co-hosts. Carrie communicated this idea to Astrid in their recorded conversation:

(62) There's this idea that if you come out particularly in later life, you're gonna be run out of town you know with yokels with flaming torches and stuff

Carrie also echoes this idea in her conversation with Astrid when she discusses encountering people she knows in real life talking about trans women on social media:

(63) There's lots of people that I know, personally and professionally, who it turns out think I'm a monster, and will happily say that people like me are monsters - in the most polite, middle class, dinner party kind of way

Carrie's use of harsh voice could be connected to the harsh voice used in portrayals of working-class Scottish masculinity such as Rab C. Nesbitt, as discussed by Stuart-Smith (1999b). I also connect this to a common trope in TV and film representations of trans women and cisgender men who have disguised themselves as cisgender women being revealed to be male. For example, in *The Boxtrolls* (2014), the villain, Snatcher, disguises himself as a woman named Madame Frou-Frou. Upon being revealed, Snatcher imitates his high-pitched Madame Frou-Frou voice before switching to back to harsh voice, creating dissonance between the two characters. The two voices are accompanied by shifts in accent features, with Snatcher using Cockney features such as th-fronting and glottal stops in his own voice and a German accent as Madame Frou-Frou. Another example comes from the slasher film *Sleepaway Camp* (1983). At the end of the film, the character Angela is revealed be both her dead brother and the murderer. In the scene depicted in Figure 13.2, she stands nude and covered in blood, she makes use of harsh voice and whisper as she lets out an animalistic growl that lacks lexical content.

Figure 13.2: Angela is revealed to be Peter. From Sleepaway Camp (1983). Image source: https://indiemacuser.com/2015/06/05/horror-month-interview-with-felissa-rose-sleepaway-camp/

Carrie's use of harsh voice mirrors these portrayals. When using harsh voice to compare her voice to that of her cisgender co-host, she also shifts her accent features towards Cockney, using h-dropping ("ello!'), while other instances, such as the example from the open mic story, lack lexical content. I therefore suggest that Carrie draws from these media representations of men disguising themselves as women in her linguistic behaviour to convey how she understands other to perceive her voice, and that these media portrayals ultimately impact how she sees her own voice. Her capacity to exploit harsh voice in this way suggests evidence of an iconized link between harsh voice and trans womanhood that circulates through cultural depictions of trans women as monstrous, villainous and deceptive.

13.4 Self-acceptance as the way forward

Carrie frames self-acceptance as being the way forward, and that despite her seeing limits to self-acceptance, she sees self-acceptance as having power on both a personal level and as a driver of wider societal change.

Carrie's experience of time can be framed in the lens of trans temporality, a concept that R. Pearce (2018: 119–156) explores to describe the way that trans individuals' emotional engagement of time is shaped by their experiences. She draws on Bradley & Myerscough's (2015) 'time of anticipation' and Halberstam's (2005) concept of queer time to discuss how lengthy waiting lists and anticipation of difficulty and mistrust in interactions with gender clinics affect how trans people engage with the experinec of time. Carrie's experience with waiting lists plays a role in the sense of agency she feels she has with her voice. She mentions being about to go on a waiting list for gender confirmation surgery and contemplates 'maybe when I've been through that I might revisit the voice thing', but notes that 'I've got a few years still to go'. R. Pearce (2018: 124) notes that 'we can regard the transitioning body as simultaneously rooted in a future through anticipation [...] and in the past through social readings that sex the body's physical frame'. This is echoed in Carrie's presentation of the future as involving both societal acceptance of trans people and self-acceptance of her own voice, while framing the past as dangerous and stifling.

Trans temporality is useful for exploring the role that Carrie's fear of the 'world where I'm terrified everybody's gonna kill me' shapes her vocal behaviour, in spite of feeling that there is 'no evidence' for it. R. Pearce (2018: 132) explores how patients waiting for gender clinic appointments anticipate experience mistreatment from health professionals on the basis of communally constructed accounts of transphobia. In this way, the 'real or imagined past of another' becomes implicated in trans people's anticipation of the future as they wait for appointments. Similarly, Carrie describes

how 'years of being in the closet and being scared to come out' have left her scared of other people's reactions. In this way, Carrie's engagement with the fear she felt in her past continues to shape her behaviour, leading her to avoid presenting as female around her children and ultimately shaping her vocal practice, too.

Carrie shows increasing resistance to judgements of others, however, instead choosing to focus on a future where self-acceptance plays a pivotal role. Her account of this and how these feelings affect what she does with her voice mirrors the experiences of Zimman's (2017a) transmasculine participants in the early stages of testosterone therapy. Zimman explores how as participants' voices lowered and they began to be gendered as male by others, some participants felt more comfortable incorporating less traditionally masculine characteristics into their presentation as well as their vocal style. Though starting feminising hormone therapy does not have an effect on the voice, Carrie describes a similar process. When she first came out, she says 'it was all skirts and too much makeup', and while she still feels aware of not looking the way that she wants to without wearing a wig and makeup, she feels less pressure to conform in other respects ('I can't think of the last time I got the epilator out to do my legs') and no longer feels she has to be 'hyper-femme' to prove her gender identity to others. She is also increasingly resistant to changing her voice for others ('Just because you think this is how a woman should talk - I'm not gonna talk like that because I am still me'), something she embraces in her music in particular ('I'm a trans woman singing in a rock band, that's what we sound like! [...] and it's a good thing'). Zimman's (2017a) work focused on the relationship between f0 and /s/, and he interprets similar findings in his work (that some participants use a higher COG /s/ as their voice pitch drops) through a lens of stylistic bricolage, arguing that being gendered as male by others and speaking with a lower pitched voice can re-contextualise the meaning of a higher-frequency /s/ from being interpreted as cisnormative femininity to indexing a gay or queer male identity.

The concept of stylistic bricolage (Hebdige 1979, Eckert 2004) is potentially relevant here. Carrie's experiences as a trans woman, however, differ from the experiences of transmasculine participants in Zimman's study in ways that influence her decisions about her presentation, and in turn, her vocal behaviour. Unlike for Zimman's (2017a) transmasculine participants, for whom the physiological effects of hormones play a major role in pitch change, increasing pitch for Carrie involves physical effort and commitment. In the later interviews with Zimman's participants, this pitch decrease is part of a larger change to their appearance that means that others typically gender them as male, allowing them in turn to incorporate less gender normative features into their style. Carrie, however, still reports still being misgendered on the basis of expectations about what a woman's voice should sound like. Incorporating features that would be interpreted as masculine by others while presenting female for Carrie, especially in places like changing rooms and toilets where the presence of trans women

is often contested and politicised, is associated with fear and danger. Because of these consequences, Carrie faces pressure to maintain congruence between her voice and presentation, and her ability to sustain a higher pitch voice is dependent on whether she feels it is appropriate to her wider style and context. She has no trouble using higher pitch when she sings, for example, but has trouble sustaining the voice she was learning in voice therapy, which she describes as sounding like 'a children's TV presenter'. This voice involves not only increasing pitch, but differences in intonation, as well as physical facial movement — wide eyes and a mobile face. Shifting a single stylistic feature in isolation, be that wearing a wig, increasing pitch, producing a breathier voice, or opening her eyes wider, is something that Carrie appears to find difficult. This may be partly due to how this was taught in voice therapy, as a unified style rather than a pitch increase in isolation, but for Carrie it appears to be more related to having to retain the ability to move between two worlds and the repercussions that come with using the 'man voice' voice in the wrong context. However, Carrie's perception that she shifts vocal features as a unified style may not necessarily align with her vocal performance.

Carrie's resistance to conforming to a conventionally feminine voice is pertinent when she discusses its role in accessing transition healthcare, describing how she was told by other trans women who had been to the gender clinic before her that she would need to conform to gender stereotypes in order to get access to healthcare, which included being conventionally attractive and losing weight alongside having a high pitched voice. Stone (2013) discusses this phenomenon more detail, exploring how early gender clinics, run by male doctors, took on the role of 'charm schools', evaluating candidates for gender confirmation surgery on the basis of their performance of gender and on candidates' ability to reproduce narratives of feeling they were living in the 'wrong body'. The criteria by which access to surgery was granted were then refined over the years in a recursive process where narratives of trans identity were shaped to fit diagnostic criteria, and diagnostic criteria were reinforced by candidates presenting narratives that fit. Carrie's resistance to the idea that her voice must sound a particular way echoes Stone's (2013) resistance to the wider 'wrong body' narrative. In resisting this, Carrie also resists a number of ideologies that Zimman (2012: 98-105) identifies in speech and language therapy literature, which frames trans women's voices as pathological and in need of intervention from cisgender voice therapists in order to sound normatively feminine.

13.5 Conclusion

Overall, Carrie's understanding of the amount of control she has over her voice and situation, the social pressures she feels to conform to standards of vocal femininity, and

her desire to resist these pressures and focus on self-acceptance, all play a role in how Carrie uses her voice to navigate her life. This study has particularly highlighted the role of wider societal transphobia and the way that trans people use their voices. While this has been noted as a potential factor in previous research, IPA allowed detailed consideration of how Carrie's understanding of her own voice paralleled transphobic media representations that frame trans women's voices as evidence of their exclusion from womanhood and position trans women as monstrous and deceptive. Furthermore, Carrie's perception that there are parts of her voice that she is unable to change, and her resistance to pressure to conform to ideas of what female voices should sound like reveals the role of how someone constructs their own agency as a factor that shapes trans people's voices, something which Zimman (2021: 85) identified as a potential avenue for future research.

13.6 A place for Interpretative Phenomenological Analysis in sociophonetic research?

The framework that IPA provided for each stage of the analysis process, from close line-by-line initial noting, to identifying themes and connections between them, was particularly useful in the present study. IPA's idiographic focus allowed consideration of how Carrie's day-to-day interactions took on particular significance for her in the context of her experience coming out and living as a trans woman. This in turn allowed her own vocal behaviour to understood in the context of how she understands her voice, how this shapes her interactions with others, and how this relates to wider societal attitudes towards trans women.

I would urge future research on trans speakers to consider the use of IPA as a potential approach. Given that sociophonetic research on trans speakers is still relatively new, I believe that its in-depth qualitative focus is useful as a way of incorporating participants' expertise on their own experience into research.

While small sample size and idiographic focus of IPA mean it is note appropriate for all research questions, it may be apt for future linguistic research where participants' understanding of a major life experience might effect on their linguistic practice. It presents a viable alternative to ethnography in cases where the focus is on individuals rather than community norms. Through focusing on the particular, IPA may help to reveal cases where individual factors might contribute to speakers deviating from community norms, in turn deepening our understanding of how macro-level speech patterns and social norms relate to variation at the level of individual speakers. Research on experiences of migration, language attrition, and speech disorder could be potential avenues - situations where an individual might be highly aware of their own speech

and the social repercussions of it.

Part IV

Conclusion

In this thesis, I considered the social meaning of laryngeal voice quality in Scottish accents. I was interested in the relationships between small-scale and large-scale aspects of voice quality and meaning, concerning myself with how detailed auditory-perceptual analysis relates to larger-scale corpus acoustic analysis, and how a single speaker's understanding of her voice and use of it in interaction related to the wider social meaning attributed to voices.

To investigate these issues, I combined auditory-perceptual and acoustic analysis first in a corpus investigation of how Scottish voice quality varied according to social and linguistic factors in speakers sampled from SCOSYA, and then in a case study of the voice of Carrie, a transgender woman, using Interpretative Phenomenological Analysis.

The corpus analysis began with auditory-perceptual analysis (Chapter 6), which used a novel VPA-inspired method termed Phonation Profile Analysis (PPA). This method revealed that Scottish voice quality is characterised by a whispery and tense, whispery quality, and Scottish voice quality varied by area, with more tense, whispery voice in Glasgow, more breathy voice in Lothian and more creaky voice in Shetland.

I connected this to acoustic analysis of voice quality, first considering how PPA related to an automated f0-based method of identifying creak (Chapter 7), and then how non-creaky voice quality could be characterised through a multi-measure analysis (Chapter 8). The f0-based method operated at a different time-scale from PPA, and faced issues with certain speakers and voice qualities. The multi-measure acoustic analysis showed that whispery voice, breathy voice, tense voice, modal voice, and tense whispery voice varied in their acoustic manifestation, but found that there was still considerable variation within each phonation type and that multiple measures were necessary to characterise variation.

The larger-scale corpus study (Chapter 9) confirmed some of the findings of the smaller-scale PPA, such as the prevalence of creak among Insular Scots speakers, and revealed new insights such as variation according to age, gender and linguistic factors that cut across the use of distinct phonation types: Female speaker's voices displayed tenser voice quality that male speakers overall, while younger speakers' voices were tenser and more modal-like than older speakers.

In the case study of Carrie's voice (Part III), Interpretative Phenomenological Analysis revealed three overarching themes that explored how Carrie's understanding of her experiences with her voice related to how she uses her voice in social interactions and how her voice is attributed social meaning by others. The quantitative acoustic analysis revealed that Carrie's voice quality may range between breathy and whispery, developing a laxer, more modal-like quality in conversation with an unknown interlocutor. Qualitative acoustic analysis revealed that she made use of a harsh 'beast' voice to

convey how she understands others to view her voice as monstrous.

Taken together, these analyses have important implications for the study of voice quality and social meaning moving forward. Firstly, in both the corpus study and the case study, auditory-perceptual analysis allowed acoustic data to become meaningful. Together, the investigation of how PPA and acoustic measurement related to each other presented in Chapter 7 and Chapter 8 demonstrated that acoustic measurement cannot be separated from other domains of voice quality (Kreiman & Gerratt 2018), and that a number of acoustic measures need to be used together in order to characterise variation in voices (Kreiman et al. 2021).

The implementation of these methods presented challenges and limitations, and I suggest that future research refines application of these methods. Firstly, PPA requires the researcher to distinguish between such a large number of phonation types and scalar degrees may reduce its usability. I suggest, in future research, it may be useful to develop PPA following Segundo & Mompean (2017) simplification of VPA, and aim to find the smallest number of parameters necessary to characterise a speaker's phonation profile. I also find limitations in the unit of the 'Voiced Stretch' in terms of how the variation in its length and segmental composition likely affects perceived quality. I suggest further research should consider alternative units of analysis, and investigate the effect of unit of analysis on intra- and inter-rater reliability.

In Chapter 7 and Chapter 9, I used f0-based identification of creaky voice to separate creaky and non-creaky voice. This allowed the characteristic creaky quality of Insular speakers to be confirmed acoustically and a clearer analysis of non-creaky voice quality, revealing the tenser quality of female speakers and the more modal-like quality of younger speakers. However, I also find a number of limitations in this approach, particularly in the decreased performance in older speakers and those whose creaky and non-creaky f0s overlap considerably, such as Carrie. I would suggest that future research turns toward the Union method for identifying creak presented by White et al. (2022), rather than f0-based analysis alone. Future research should draw on the analysis presented in Chapter 7 to consider cases where the Union method and manual coding of creak do not align in more detail, which might reveal areas for improvement in future adaptations.

I also suggest that the auditory-acoustic approach taken to characterising Carrie's voice quality could be refined further for future research. This approach developed out of the fact that Carrie's use of creak could not be identified using the automated f0-based approach, but came to reveal interactional and linguistic influences of creak, whisper and harsh voice in a way that an automated approach would have missed. Carrie's use of harsh voice, harsh whispery voice and laryngeal constriction at high pitch for different semiotic purposes also emphasises the role of the laryngeal constrictor

in voice quality (Esling et al. 2019).

Throughout this thesis, the influence of linguistic factors on the voice has also been apparent. Chapter 9 showed that duration, speech rate, glottalisation, aspiration, pre-aspiration phrase position, phrase-initial vowels, and the consonants contained in a stretch all influence voice quality. While the influence of many of these factors is well-established, their influence raises a key question for future research: Where do segmental features end, and where does voice quality begin?

Furthermore, when the findings of the corpus study in terms of gender variation are taken alongside the case study of Carrie's voice, they reveal the importance of context in the creation of social meaning. Female speakers with Scottish accents show a tenser voice quality than male speakers, and demonstrate that associations between breathy voice and female speakers are culturally-specific rather than universal. Carrie, however, shows awareness of the wider link between breathy voice and female speakers as she aims to produce 'a more breathy voice than God intended', showing that understandings of breathy voice as a feature of female voices persist even while Scottish female speakers in the general population show a tenser voice quality than male speakers. Future research should consider the perception of voice quality in Scottish accents in light of this, and consider how tenser and breathier voice qualities are perceived in the context of Scottish and non-Scottish accents.

In the same way that social meaning of voice quality emerges in the context of other variables and wider ideologies about voice's speaker, the meaning of a particular acoustic measure cannot be removed from the context of other measures and the auditory quality of a voice. Taken together, the distinct parts of this study show that to consider the meaning of a voice, we must take the voice not in part, but as a whole, and consider its auditory quality, its acoustic form, and social meaning together.

Bibliography

Abdelli-Beruh, Nassima B, Lesley Wolk & Dianne Slavin. 2014. Prevalence of vocal fry in young adult male American English speakers. *Journal of Voice* 28(2). 185–190.

Abercrombie, David. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Agha, Asif. 2005. Voice, footing, enregisterment. *Journal of Linguistic Anthropology* 15(1). 38–59.

de Araújo Pernambuco, Leandro, Albert Espelt, Patrícia Maria Mendes Balata & Kenio Costa de Lima. 2015. Prevalence of voice disorders in the elderly: a systematic review of population-based studies. *European Archives of Oto-Rhino-Laryngology* 272. 2601–2609.

Babel, Anna M. 2016. Silence as control: shame and self-consciousness in sociolinguistic positioning. In Anna M. Babel (ed.), 200–227. Cambridge University Press.

Ball, Martin J., John Esling & Craig Dickson. 1995. The VoQS system for the transcription of voice quality. *Journal of the International Phonetic Association* 25(2). 71–80.

Barnes, Michael. 1984. Orkney and Shetland. In *Language in the British Isles*, 352–366. Cambridge: Cambridge University Press.

Barnes, Michael. 1991. Reflections of the structure and the demise of Orkney and Shetland Norn. In P Sture Ureland & George Broderick (eds.), *Language contact in the British Isles: Proceedings of the Eighth International Symposium on Language Contact in Europe, Douglas, Isle of Man, 1988*, 429–460. Niemeyer.

Barras, Claude. 2002. *Transcriber*. Version 1.5.1.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Beck, Janet. 1988. *Organic variation and voice quality*. University of Edinburgh dissertation.

Beck, Janet. 2005. Perceptual analysis of voice quality: the place of the Vocal Profile Analysis. In Janet Beck & William J Hardcastle (eds.), *A figure of speech: a festschrift for John Laver*, 232–285. Mahwah, New Jersey: Lawrence Erlbaum Associates.

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

Becker, Kara, Sameer ud Dowla Khan & Lal Zimman. 2022. Beyond binary gender: creaky voice, gender, and the variationist enterprise. *Language Variation and Change* 34(2). 215–238.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204.

Belsley, D A, E Kuh & R E Welsch. 1980. *Regression diagnostics. identifying influential data and sources of collinearity*. New York: Wiley.

van den Berg, Janwillem. 1958. Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research* 1(3). 227–244.

van den Berg, Janwillem. 1968. Mechanism of the larynx and the laryngeal vibrations. In Bertil Malmberg (ed.), *Manual of phonetics*. Amsterdam; London: North-Holland Pub. Co.

Bickley, Corinne. 1982. Acoustic analysis and perception of breathy vowels. 1. 73–83.

Bishop, Jason & Patricia Keating. 2012. Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex. *The Journal of the Acoustical Society of America* 132. 1100–1112.

Boersma, Paul. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proceedings 17* 17. 97–110.

Bogert, Bruce P. 1963. The quefrency alanysis of time series for echoes: cepstrum, pseudoautocovariance, cross-cepstrum and saphe cracking. In *Proc. symposium time series analysis, 1963*, 209–243.

Bois, John W. Du. 2007. The stance triangle. In R. Englebretson (ed.), *Stancetaking in discourse: subjectivity, evaluation, interaction*, 139–182. Amsterdam & Philadelphia: John Benjamins Publishing.

Bottalico, Pasquale, Juliana Codino, Lady Catherine Cantor-Cutiva, Katherine Marks, Charles J. Nudelman, Jean Skeffington, Rahul Shrivastav, Maria Cristina Jackson-Menaldi, Eric J. Hunter & Adam D. Rubin. 2018. Reproducibility of voice parameters: the effect of room acoustics and microphones. *Journal of Voice* 34(3). 320–334.

Boyd, Zac & Míša Hejná. 2022. Friend or foe? the 'Critical Role' of voice quality in Dungeons & Dragons via non-player characters voiced by Matthew Mercer. In Stanford, USA.

Bradley, Jess & Francis Myerscough. 2015. *Transitional demands*. `http://actionfortransheal org.uk/2015/03/30/transitional-demands/`.

Camacho, Arturo & John G. Harris. 2008. A sawtooth waveform inspired pitch estimator for speech and music. *The Journal of the Acoustical Society of America* 124(3). 1638–1652.

Catford, J. C. 1957. Vowel-systems of Scots dialects. *Transactions of the Philological Society* 56(1). 107–117.

Catford, J. C. 1977. *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.

Chai, Yuan & Marc Garellek. 2019. Using H1 instead of H1–H2 as an acoustic correlate of glottal constriction. *The Journal of the Acoustical Society of America* 146(4). 3008–3008.

Chasaide, Ailbhe Ní & Christer Gobl. 1993. Contextual variation of the vowel voice source as a function of adjacent consonants. *Language and Speech* 36(2-3). 303–330.

Chen, Gang, Xue Feng, Yen-Liang Shue & Abeer Alwan. 2010. On using voice source measures in automatic gender classification of children's speech. In *Proc. interspeech 2010*, 673–676.

Coleman, E, AE Radix, WP Bouman, GR Brown, ALC De Vries, MB Deutsch, R Ettner, L Fraser, M Goodman, J Green, et al. 2022. Standards of care for the health of transgender and gender diverse people, version 8. *International Journal of Transgender Health* 23(sup1). S1–S259.

Creese, Angela & Fiona Copland. 2015. *Linguistic ethnography: collecting, analysing and presenting data*. SAGE Publications Ltd.

Creswell, John W. 2015. *A concise introduction to mixed methods research*. Thousand Oaks, CA: Sage.

Creswell, John W. 2017. *Research design: qualitative, quantitative, and mixed methods approaches*. 5th. London: SAGE Publications.

Croissant, Yves. 2020. Estimation of random utility models in r: the mlogit package. *Journal of Statistical Software* 95(11). 1–41.

Crowley, Archie. 2021. The weight of the voice: gender, privilege, and qualic apperception. *Toronto Working Papers in Linguistics* 43(1).

Cruttenden, Alan & A. C. Gimson. 2014. *Gimson's pronunciation of English*. English. Eighth. London: Routledge.

D'Onofrio, Annette & Penelope Eckert. 2021. Affect and iconicity in phonological variation. *Language in Society* 50(1). 29–51.

Dahl, David B, David Scott, Charles Roosen, Arni Magnusson & Jonathan Swinton. 2019. *xtable: Export Tables to LaTeX or HTML*.

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

Dallaston, Katherine & Gerard Docherty. 2020. The quantitative prevalence of creaky voice (vocal fry) in varieties of english: a systematic review of the literature. *PLOS ONE* 15(3). 1–18.

Daly, Mary. 1978. *Gyn/ecology: the metaethics of radical feminism*. Boston: Beacon.

Dave, R. V. 1977. *Studies in Gujarati phonology and phonetics*. Cornell University dissertation.

Davies, Shelagh, Viktória G. Papp & Christella Antoni. 2015. Voice and communication change for gender nonconforming individuals: giving voice to the person inside. *International Journal of Transgenderism* 16. 117–159.

De Krom, Guus. 1995. Spectral Correlates of Breathiness and Roughness for Different Types of Vowel Fragments. *Journal of Speech, Language, and Hearing Research* 38(October). 794–811.

Deevi, Sathish & 4D Strategies. 2016. *modes: Find the Modes and Assess the Modality of Complex and Mixture Distributions, Especially with Big Datasets*.

Deliyski, Dimitar D., Maegan K. Evans & Heather S. Shaw. 2005. Influence of data acquisition environment on accuracy of acoustic voice quality measurements. *Journal of Voice* 19(2). 176–186.

Deliyski, Dimitar D., Heather S. Shaw & Maegan K. Evans. 2005. Adverse effects of environmental noise on acoustic voice quality measurements. *Journal of Voice* 19(1). 15–28.

Dilley, Laura, Stefanie Shattuck-Hufnagel & Mari Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of phonetics* 24(4). 423–444.

Docherty, Gerry & Paul Foulkes. 1999. Sociophonetic variation in 'glottals' in Newcastle English. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.), *Proceedings of 14th International Congress of Phonetic Sciences, San Francisco, CA, USA*, 1037–1040.

Dorreen, Keiran. 2017. *Fundamental frequency distributions of bilingual speakers in forensic speaker comparison*. University of Canterbury MA thesis.

ud Dowla Khan, Sameer. 2012. The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics* 40(6). 780–795.

Eckert, Penelope. 2004. The meaning of style. *Texas Linguistic Forum* 47. 41–43.

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

Eckert, Penelope. 2012. Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology* 41. 87–100.

Eckert, Penelope. 2017. Comment: the most perfect of signs: iconicity in variation. *Linguistics* 55(5). 1197–1207.

Eckert, Penelope & William Labov. 2017. Phonetics, phonology and social meaning. *Journal of Sociolinguistics* 21(4). 467–496.

Edmondson, Jerold A & John Esling. 2006. The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology* 23(2). 157–191.

Eisenberg, Elodie & Karyofyllis Zervoulis. 2020. All flowers bloom differently: an interpretative phenomenological analysis of the experiences of adult transgender women. *Psychology & Sexuality* 11(1-2). 120–134.

Elyan, Olwen, Philip Smith, Howard Giles & Richard Bourhis. 1978. Rp-accented female speech: the voice of perceived androgyny. *Sociolinguistic patterns in British English*. 122–131.

English, Kirstie Ken. 2022. Binary barriers: representing trans & gender diverse populations in censuses & other population surveys. In *Mind the gap conference july 2022*.

Epstein, Melissa. 2000. A comparison of linguistic and pathological breathiness using the LF model. *The Journal of the Acoustical Society of America* 107(5). 2906–2906.

Epstein, Melissa. 2002. Voice Quality and Prosody in English. *Time* (August). 2405–2408.

Erikson, F. 1986. Qualitative methods in research on teaching. In M. C. Wittrock (ed.), 119–158. New York: Macmillan.

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

Esling, John. 1978b. *Voice quality in Edinburgh: A Sociolinguistic and Phonetic Study*. University of Edinburgh dissertation.

Esling, John. 2005. There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics* 50(1-4). 13–44.

Esling, John, Scott Moisik, Allison Benner & Lise Crevier-Buchman. 2019. *Voice Quality: The Laryngeal Articulator Model* (Cambridge Studies in Linguistics). Cambridge University Press.

Esposito, Christina M. 2010. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* 38(2). 306–316.

Esposito, Christina M. & Sameer ud Dowla Khan. 2020. The cross-linguistic patterns of phonation types. *Language and Linguistics Compass* 14(12). e12392.

Fant, Gunnar. 1960. *Acoustic theory of speech production*. The Hague: Mouton.

Fant, Gunnar, Johan Liljencrants, Qi-guang Lin, et al. 1985. A four-parameter model of glottal flow. *STL-QPSR* 4(1985). 1–13.

Farr, Joanna, Laura Blenkiron, Richard Harris & Jonathan A Smith. 2018. "it's my language, my culture, and it's personal!" migrant mothers' experience of language use and identity change in their relationship with their children: an interpretative phenomenological analysis. *Journal of Family Issues* 39(11). 3029–3054.

Firke, Sam. 2020. *janitor: Simple Tools for Examining and Cleaning Dirty Data*.

Fischer-Jørgensen, Eli. 1967. Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics* 28. 71–139.

Flanagan, J. L. 1957. Note on the design of "terminal-analog" speech synthesizers. *The Journal of the Acoustical Society of America* 29. 306–310.

Flanagan, J. L. 1958. Some properties of the sound source. *Journal of speech and Hearing research* 1. 99–116.

Frankenstein. 1931. Dir. James Whale.

Fromont, Robert. 2022. *trs2grid.jar*. `https://sourceforge.net/projects/labbcat/`.

Garellek, Marc. 2012. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics* 40(1). 152–161.

Garellek, Marc. 2015. Perception of glottalization and phrase-final creak. *The Journal of the Acoustical Society of America* 137(2). 822–831.

Garellek, Marc. 2019. The phonetics of voice. In *The routledge handbook of phonetics*, 75–106.

Garellek, Marc, Patricia Keating, Christina M. Esposito & Jody Kreiman. 2013. Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America* 133(2). 1078–1089.

Garellek, Marc, Amanda Ritchart & Jianjing Kuang. 2016. Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics* 59. 110–121.

Garellek, Marc, Robin Samlan, Bruce Gerratt & Jody Kreiman. 2016. Modeling the voice source in terms of spectral slopes. *The Journal of the Acoustical Society of America* 139(3). 1404–1410.

Gerratt, Bruce & Jody Kreiman. 2001. Toward a taxonomy of nonmodal phonation. *Journal of Phonetics* 29. 365–381.

Gerratt, Bruce, Jody Kreiman, Norma Antonanzas-Barroso & Gerald S Berke. 1993. Comparing internal and external standards in voice quality judgments. *Journal of Speech, Language, and Hearing Research* 36(1). 14–20.

Gerratt, Bruce, Jody Kreiman & Marc Garellek. 2016. Comparing measures of voice quality from sustained phonation and continuous speech. *Journal of Speech, Language, and Hearing Research* 59(5). 994–1001.

Gibson, Andy & Allan Bell. 2012. Popular music singing as referee design. *Style-Shifting in Public. New Perspectives on Stylistic Variation*. 139–164.

Gick, Bryan, Ian Wilson, Karsten Koch & Clare Cook. 2004. Language-specific articulatory settings: evidence from inter-utterance rest position. *Phonetica* 61(4). 220–233.

Gittelson, Ben, Adrian Leemann & Fabian Tomaschek. 2021. Using Crowd-Sourced Speech Data to Study Socially Constrained Variation in Nonmodal Phonation. *Frontiers in Artificial Intelligence* 3(January). 1–9.

Gobl, Christer. 1989. A preliminary study of acoustic voice quality correlates. *STL-QPSR* 4(9-21). 534.

Gobl, Christer & Ailbhe Ní Chasaide. 1992. Acoustic characteristics of voice quality. *Speech Communication* 11. 481–490.

Gobl, Christer & Ailbhe Ní Chasaide. 2000. Testing affective correlates of voice quality through analysis and resynthesis. In *Isca tutorial and research workshop (itrw) on speech and emotion*.

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

Gordeeva, Olga B. & James M. Scobbie. 2010. Preaspiration as a correlate of word-final voice in Scottish English fricatives. English. In Susanne Fuchs, Martine Toda, Marzena Żygis & ProQuest (Firm) (eds.), *Turbulent sounds: an interdisciplinary guide*, vol. 21;21.; 167–207. Berlin;New York, N.Y; Mouton de Gruyter.

Gordeeva, Olga B. & James M. Scobbie. 2013. A phonetically versatile contrast: pulmonic and glottalic voicelessness in Scottish English obstruents and voice quality. *Journal of the International Phonetic Association* 43(3). 249–271.

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

Gray, Henry. 1918. *Anatomy of the human body*. Phildelphia & New York: Lea & Febiger.

Gregersen, F, S B Nielsen & J Thøgersen. 2009. Steeping into the same river twice: on the discourse context analysis in the LANCHART project. *Linguistica Hafniensia* 41. 30–63.

Gumperz, John J. 1982. *Discourse strategies* (Studies in Interactional Sociolinguistics). Cambridge University Press.

Gussenhoven, Carlos. 2002. Intonation and interpretation: phonetics and phonology. In *Speech prosody 2002, international conference*.

Halberstam, Jack. 2005. *In a queer time and place: transgender bodies, subcultural lives*. Vol. 3. NYU press.

Hall-Lew, Lauren, Emma Moore & Robert Podesva. 2021. Social meaning and linguistic variation: theoretical foundations. In Lauren Hall-Lew, Emma Moore & Robert Podesva (eds.), *Social meaning and linguistic variation: theorizing the third wave*, 1–24. Cambridge University Press.

Hannaford, Philip C, Julie A Simpson, Ann Fiona Bissetand, Adrian Davis, William McKerrow & Robert Mills. 2005. The prevalence of ear, nose and throat problems in the community: results from a national cross-sectional postal survey in Scotland. *Family Practice* 22(3). 227–233.

Hanson, Helen M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America* 101(1). 466–481.

Hanson, Helen M. & E.S Chuang. 1999. Glottal characteristics of male speakers: acoustic correlates and comparisons with female data. *Journal of Acoustical Society of America* 106. 1064–1077.

Harrison, Nicola, Lisa Jacobs & Adrian Parke. 2020. Understanding the lived experiences of transitioning adults with gender dysphoria in the united kingdom: an interpretative phenomenological analysis. *Journal of LGBT Issues in Counseling* 14(1). 38–55.

Hartig, Florian. 2022. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. R package version 0.4.6.

Hartmann, Graham. 2019. 15 'ugly' singing voices that rule. *Loudwire*.

Hausman, J A & D McFadden. 1984. A Specification Test for the Multinomial Logit Model. *Econometrica* (52). 1124–1219.

Hazenberg, Evan. 2016. Walking the straight and narrow: linguistic choice and gendered presentation. *Gender and Language* 10(2). 270–294.

Hebdige, Dick. 1979. *Subculture: the meaning of style*. New York: Routledge.

Hejná, Míša. 2015. *Pre-aspiration in Welsh English: A case study of Aberystwyth*. University of Manchester dissertation.

Hejná, Míša. 2019. A case study of menstrual cycle effects: Global phonation or also local phonatory phenomena? In Marija Tabain Sasha Calhoun Paola Escudero & Paul Warren. (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 2630–2634. Canberra, Australia: Australasian Speech Science & Technology Association Inc.

Hejná, Míša & Anna Jespersen. 2021. The coming of age: how do linguists tease apart chronological, biological and social age? *Language and Linguistics Compass* 15(1). e12404.

Hejná, Míša & Anna Jespersen. 2022. Ageing well: social but also biological reasons for age-grading. *Language and Linguistics Compass* 16(5-6). e12450.

Hejná, Míša & Jane Scanlon. 2015. New laryngeal allophony in Manchester English. In *International Congress of Phonetic Sciences (ICPhS 2015), Glasgow*.

Hejná, Míša, Pavel Šturm, Lea Tylečková & Tomáš Bořil. 2021. Normophonic breathiness in Czech and Danish: are females breathier than males? *Journal of Voice* 35(3). 498.e1–498.e22.

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

Henton, Caroline & Anthony Bladon. 1988. Creak as a sociophonetic marker. In L M Hyman & C N Li (eds.), *Language, speech and mind: studies in honour of victoria a. fromkin*, 3–29.

Hillenbrand, James, RA Cleveland & RL Erickson. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech, Language and Hearing Research* 37. 769–778.

Hillenbrand, James & Robert A. Houde. 1996. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research* 39(2). 311–321.

Hines, Sally. 2020. Sex wars and (trans) gender panics: identity and body politics in contemporary uk feminism. *The Sociological Review* 68(4). 699–717.

Hirano, M. 1981. *Clinical examination of voice*. Vol. 5. New York;Wien; Springer-Verlag.

Hollien, Harry. 1974. On vocal registers. *Journal of Phonetics* 2. 125–143.

Hollien, Harry, Paul Moore, Ronald W Wendahl & John F Michel. 1966. On the nature of vocal fry. *Journal of Speech and Hearing Research* 9. 245–247.

Holmberg, Eva B., Robert E. Hillman, Joseph S. Perkell, Peter C. Guiod & Susan L. Goldman. 1995. Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech, Language, and Hearing Research* 38. 1212–1223.

Honikman, Beatrice. 1964. Articulatory settings. In *In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday, 12 September 1961*. London: Longmans, Green & Co. ltd.

Hymes, Dell. 1968. The ethnography of speaking. In Joshua A. Fishman (ed.), *Readings in the Sociology of Language*, 99–138. Berlin, Boston: De Gruyter Mouton.

Hymes, Dell. 1972. On communicative competence. In J.B. Pride & J. Holmes (eds.), vol. 269293, 269–293. Harmondsworth: Penguin.

Irvine, Judith & Susan Gal. 2000. Language ideology and linguistic differentiation. In Paul V. Kroskrity (ed.), *Regimes of language: ideologies, politics, and identities*, 35–83. Santa Fe, NM: SAR Press.

Iseli, Markus & Abeer Alwan. 2004. An improved correction formula for the estimation of voice source harmonic magnitudes. *The Journal of the Acoustical Society of America* 115(5). 2610–2610.

Iseli, Markus, Yen-Liang Shue & Abeer Alwan. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice sourcea). *The Journal of the Acoustical Society of America* 121(4). 2283–2295.

Ishi, Carlos Toshinori, Ken-Ichi Sakakibara, Hiroshi Ishiguro & Norihiro Hagita. 2008. A method for automatic detection of vocal fry. *IEEE Transactions on Audio, Speech, and Language Processing* 16(1). 47–56.

Izzard, Suzy Eddie. 2023. *As people may now well know, I have added the name 'Suzy' to my names*. https://twitter.com/eddieizzard/status/1664240649787559936.

Jeong, Sunwoo. 2017. Iconization of sociolinguistic variables: the case of archetypal female characters in classic hollywood cinema. English. In Matthias Bauer (ed.), *Dimensions of iconicity*, vol. 15. Philadelphia, [Pennsylvania];Amsterdam, [Netherlands]; John Benjamins Publishing Company.

Johnston, Paul. 1997. Regional Variation. In Charles Jones (ed.), *The Edinburgh history of the Scots language*, 433–513. Edinburgh: Edinburgh University Press.

Johnstone, Barbara & Scott F. Kiesling. 2008. Indexicality and experience: Exploring the meanings of /aw/-monophthongization in Pittsburgh. *Journal of Sociolinguistics* 12(1). 5–33.

Jones, Jacq. 2022. *Authentic self, incongruent acoustics: a corpus-based sociophonetic analysis of nonbinary speech*. University of Canterbury dissertation.

Kane, John, Thomas Drugman & Christer Gobl. 2013. Improved automatic detection of creak. *Computer Speech & Language* 27(4). 1028–1047.

Kawahara, H, M Morise, T Takahashi, R Nisimura, T Irino & H Banno. 2008. TandemSTRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In

2008 ieee international conference on acoustics, speech and signal processing, 3933–3936.

Kawahara, Hideki, Hideki Banno Alain de Cheveigne, Toru Takahashi & Toshio Irino. 2005. Nearly defect-free f0 trajectory extraction for expressive speech modifications based on straight. In 537–540.

Keating, Patricia, Christina M. Esposito, Marc Garellek, Sameer ud Dowla Khan & Jianjing Kuang. 2010. Phonation contrasts across languages. *UCLA Working Papers in Phonetics* 108.

Keating, Patricia, Jianjing Kuang andMarc Garellek, Jianjing Kuang & Sameer Khan. 2021. Cross-language acoustic space for vocalic phonation distinctions. *[In preparation]*.

Keating, Patricia, Marc Garellek & Jody Kreiman. 2015. Acoustic properties of different kinds of creaky voice. In *Acoustic properties of different kinds of creaky voice*.

Kendall, Katherine. 2007. Presbyphonia: a review. *Current Opinion in Otolaryngology & Head and Neck Surgery* 15.

King, Daniel, Carrie Paechter & Maranda Ridgway. 2020. *Reform of the Gender Recognition Act: Analysis of Consultation Responses*. `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/919890/Analysis_of_responses_Gender_Recognition_Act.pdf`.

Klatt, Dennis & Laura Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America* 87. 820–857.

Knowles, Gerald. 1973. *Scouse: the urban dialect of liverpool*. University of Leeds dissertation.

Kojima, H, WJ Gould, A Lambiase & N Isshiki. 1980. Computer analysis of hoarseness. 89. 547–554.

Kowalchuk, Talia. 2020. *What trans* folks talk about when they talk about voice: learning about voice feminization on reddit*. Dalhousie University MA thesis.

Kreiman, Jody. 1982. Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics* 10(2). 163–175.

Kreiman, Jody, Norma Antoñanzas-Barroso & Bruce Gerratt. 2010. Integrated software for analysis and synthesis of voice quality. *Behavior Research Methods* 42. 1030–1041.

Kreiman, Jody & Bruce Gerratt. 1996. The perceptual structure of pathologic voice quality. *The Journal of the Acoustical Society of America* 100(3). 1787–1795.

Kreiman, Jody & Bruce Gerratt. 2000. Sources of listener disagreement in voice quality assessment. *The Journal of the Acoustical Society of America* 108(4). 1867–1876.

Kreiman, Jody & Bruce Gerratt. 2005. Perception of aperiodicity in pathological voice. *The Journal of the Acoustical Society of America* 117. 2201–2211.

Kreiman, Jody & Bruce Gerratt. 2012. Perceptual interaction of the harmonic source and noise in voice. *The Journal of the Acoustical Society of America* 131(1). 492–500.

Kreiman, Jody & Bruce Gerratt. 2018. Reconsidering the nature of voice. *The Oxford Handbook of Voice Perception* (October 2021). 166–188.

Kreiman, Jody, Bruce Gerratt & Norma Antoñanzas-Barroso. 2007. Measures of the glottal source spectrum. *Journal of Speech, Language, and Hearing Research* 50(3). 595–610.

Kreiman, Jody, Bruce Gerratt & Diana Vanlancker-Sidtis. 2003. Defining and measuring voice quality. In *Voqual'03*, 115–120.

Kreiman, Jody, Bruce Gerratt., Marc Garellek, Robin Samlan & Zhaoyan Zhang. 2014. Toward a unified theory of voice production and perception. *Loquens* 1(1). e009.

Kreiman, Jody, Yoonjeong Lee, Marc Garellek, Robin Samlan & Bruce Gerratt. 2021. Validating a psychoacoustic model of voice quality. *The Journal of the Acoustical Society of America* 149(1). 457–465.

Kreiman, Jody, Yen-Liang Shue, Gang Chen, Markus Iseli, Bruce Gerratt, Juergen Neubauer & Abeer Alwan. 2012. Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. *The Journal of the Acoustical Society of America* 132(4). 2625–2632.

de Krom, Guus. 1993. A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech, Language, and Hearing Research* 36. 254–266.

Kuang, Jianjing. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America* 142(3). 1693–1706.

Labov, William. 1963. The social motivation of a sound change. *Word* 19(3). 273–309.

Labov, William. 1984. Field methods of the project on linguistic change and variation. In John Baugh & Joel Sherzer (eds.), *Language in use*, 28–53. Engelwood Cliffs, NJ: Prentice-Hall.

Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago: University of Chicago.

Ladefoged, Peter. 2006. *A course in phonetics*. 5th edn. Boston: Thomson Wadsworth.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages*. English. Oxford: Blackwell.

Larsson, Johan & Peter Gustafsson. 2018. A case study in fitting area-proportional Euler diagrams with ellipses using eulerr. In *Proceedings of International Workshop on Set Visualization and Reasoning*, vol. 2116, 84–91. Edinburgh, United Kingdom: CEUR Workshop Proceedings.

Laver, John. 1974. Labels for voices. *Journal of the International Phonetic Association* 4(2). 62–75.

Laver, John. 1978. The concept of articulatory settings: an historical survey. *Historiographia Linguistica* 5(1-2). 1–14.

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

Laver, John. 1991. *The Gift of Speech*. Edinburgh: Edinburgh University Press.

Laver, John. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Laver, John. 1991[1979]. The description of voice quality in general phonetic theory. In *The Gift of Speech*, 184–208. Edinburgh: Edinburgh University Press.

Laver, John. 1991[1976]. The semiotic nature of phonetic data. In *The Gift of Speech*, 162–170. Edinburgh: Edinburgh University Press.

Laver, John. 1991[1968]. Voice quality and indexical information. In *The Gift of Speech*. Edinburgh: Edinburgh University Press.

Laver, John & Peter Trudgill. 1991[1979]. Phonetic and linguistic markers in speech. In *The Gift of Speech*, 184–208. Edinburgh: Edinburgh University Press.

Laver, John, Sheila Wirz, Janet Mackenzie Beck & Steven Hiller. 1991[1981]. A perceptual protocol for the analysis of vocal profiles. In *The Gift of Speech*, 265–280. Edinburgh: Edinburgh University Press.

Lee, Seung Jin, YoonHee Cho, Ji Yeon Song, DamHee Lee, Yunjung Kim & HyangHee Kim. 2016. Aging effect on korean female voice: acoustic and perceptual examinations of breathiness. *Folia Phoniatrica et Logopaedica* 67(6). 300–307.

Levon, Erez. 2007. Sexuality in context: variation and the sociolinguistic perception of identity. *Language in Society* 36. 544–54.

van Leyden, Klaske & Vincent J. van Heuven. 2006. On the prosody of Orkney and Shetland dialects. *Phonetica* 63(2-3). 149–174.

Li, K.P., G W Hughes & A S House. 1969. Correlation characteristics and dimensionality of speech spectra. *The Journal of the Acoustical Society of America* 46(4B). 1019–1025.

Lieberman, Philip. 1963. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *Journal of the Acoustical Society of America*. 344–353.

Löfqvist, Anders. 1986. The long-time-average spectrum as a tool in voice research. *Journal of Phonetics* 14(3-4). 471–475.

Lotto, A.J., L.L. Holt & K.R. Kluender. 1997. Effect of voice quality on perceived height of English vowels. *Phonetica* 54(2). 76–93.

Ḿuller, Johannes. 1837. *Handbuch der physiologie des menschen*. Vol. 2.

MacKenzie, Laurel & Danielle Turton. 2019. Assessing the accuracy of existing forced alignment software on varieties of British English. *Linguistics Vanguard*.

Maddieson, Ian. 2013. Voicing in plosives and fricatives (v2020.3). In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Zenodo.

Madill, Catherine, Duong Duy Nguyen, Kristie Yick-Ning Cham, Daniel Novakovic & Patricia McCabe. 2019. The impact of nasalance on cepstral peak prominence and harmonics-to-noise ratio. *The Laryngoscope* 129(8). E299–E304.

Martins, Regina Helena Garcia, Adriana Bueno Benito Pessin, Douglas Jorge Nassib, Anete Branco, Sergio Augusto Rodrigues & Selma Maria Michelim Matheus. 2015. Aging voice and the laryngeal muscle atrophy. *The Laryngoscope* 125(11). 2518–2521.

Maryn, Youri, Marc De Bodt & Nelson Roy. 2010. The acoustic voice quality index: toward improved treatment outcomes assessment in voice disorders. *Journal of Communication Disorders* 43(3). 161–174.

Matar, Nayla, Cristel Portes, Leonardo Lancia, Thierry Legou & Fabienne Baider. 2016. Voice quality and gender stereotypes: a study of Lebanese women with Reinke's Edema. *Journal of Speech, Language, and Hearing Research* 59(6). S1608–S1617.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger. 2017a. *Montreal Forced Aligner [Computer program]. Version 1.1.0.* `http://montrealcorpustools.github.io/Montreal-Forced-Aligner/`.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger. 2017b. Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association*.

McCarthy, Owen & Jane Stuart-Smith. 2013. Ejectives in scottish english: a social perspective. *Journal of International Phonetic Association* 43. 273–98.

Mendoza-Denton, Norma. 2011. The semiotic hitchhiker's guide to creaky voice: circulation and gendered hardcore in a Chicana/o gang persona. *Journal of Linguistic Anthropology* 21(2). 261–280.

Mills, Tyler J, Kirstie E Riddell, Elizabeth Price & David RR Smith. 2023. 'stuck in the system': an interpretative phenomenological analysis of transmasculine experiences of gender transition in the uk. *Qualitative Health Research* 33. 578–588.

Moisik, Scott R. 2013. Harsh voice quality and its association with blackness in popular American media. *Phonetica* 69(4). 193–215.

Moisik, Scott R, Míša Hejná & John H 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*, 220–224.

Moisik, Scott R. & John H. Esling. 2014. Modeling the biomechanical influence of epilaryngeal stricture on the vocal folds: a low-dimensional model of vocal–ventricular fold coupling. *Journal of Speech, Language, and Hearing Research* 57(2). S687–S704.

Monsen, Randall B. & A. Maynard Engebretson. 1977. Study of variations in the male and female glottal wave. *The Journal of the Acoustical Society of America* 62(4). 981–993.

Montiel-McCann, Camila Soledad. 2022. "it's like we are not human": discourses of humanisation and otherness in the representation of trans identity in british broadsheet newspapers. *Feminist Media Studies* 0(0). 1–17.

Moyer, Justin Wm. 2015. Lemmy dead at 70: the motörhead frontman with 'a voice like shrapnel and a bass tone to match'. *The Independent*.

Murry, Thomas & WS Brown Jr. 1971. Subglottal air pressure during two types of vocal activity: vocal fry and modal phonation. *Folia Phoniatrica et Logopaedica* 23(6). 440–449.

Nolan, Francis. 1983. *The phonetic bases of speaker recognition*. English. Cambridge: Cambridge University Press.

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

Ormston, Rachel, Liz Spencer, Matt Barnard & Dawn Snape. 2013. The Foundations of Qualitative Research. In Jane Ritchie, Jane Lewis, Carol McNaughton Nicholls & Rachel Ormston (eds.), *Qualitative research practice: a guide for social science students and researchers*, 2nd. London: SAGE Publications.

Panayotov, Vassil, Guoguo Chen, Daniel Povey & Sanjeev Khudanpur. 2015. LibriSpeech: An ASR corpus based on public domain audio books. In *IEEE International Conference on Acoustics, Speech and Signal Processing 2015 (ICASSP), Brisbane, QLD*, 5206–5210.

Paolo, Marianna Di & Alice Faber. 1990. Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language variation and change* 2(2). 155–204.

Parker, Jessica. 2020. *Changes to gender recognition laws ruled out*. https://www.bbc.co.uk/news/uk-politics-54246686.

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

Pearce, Ruth. 2018. *Understanding Trans Health: Discourse, Power and Possibility*. Bristol: Policy Press.

Pearce, Ruth, Sonja Erikainen & Ben Vincent. 2020. Terf wars: an introduction. *The Sociological Review* 68(4). 677–698.

Peirce, Charles Sanders. 1865. Five hundred and eighty-second meeting. may 14, 1867. monthly meeting; on a new list of categories. *Proceedings of the American Academy of Arts and Sciences* 7. 287–298.

Pessin, A.B.B., E.L.M. Tavares, A.C.J. Gramuglia, L.R. de Carvalho & R.H.G. Martins. 2017. Voice and ageing: clinical, endoscopic and acoustic investigation. *Clinical Otolaryngology* 42(2). 330–335.

Pharao, Nicolai, Marie Maegaard, Janus Spindler Møller & Tore Kristiansen. 2014. Indexical meanings of [s+] among Copenhagen youth: social perception of a phonetic variant in different prosodic contexts. *Language in Society* 43(1). 1–31.

Pierrehumbert, Janet & David Talkin. 1992. Lenition of /h /and glottal stop. In G. Doherty & D. R. Ladd (eds.), *Papers in laboratory phonology ii: gesture segment prosody*, 90–117. Cambridge: Cambridge University Press.

Pittam, Jeffery. 1985. *Voice quality: its measurement and functional classification*. University of Queensland dissertation.

Pittam, Jeffery. 1987. The long-term spectral measurement of voice quality as a social and personality marker: A review. *Language and Speech* 30(1). 1–12.

Pittam, Jeffery & Cynthia Gallois. 1986. Predicting impressions of speakers from voice quality: Acoustic and perceptual measures. *Journal of Language and Social Psychology* 5(4). 233–247.

Podesva, Robert. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11(4). 478–504.

Podesva, Robert. 2013. Gender and the social meaning of non-modal phonation types. In *Proceedings of the berkeley linguistics society*, vol. 37, 427–488.

Podesva, Robert. 2018. The affective roots of gender patterns in the use of creaky voice. In *Experimental and theoretical advances in prosody 4 (etap4), university of massachusetts amherst*.

Podesva, Robert J & Janneke Van Hofwegen. 2015. /s/exuality in Smalltown California: Gender Normativity and the Acoustic Realization of /s/. *Language, Sexuality, and Power: Studies in Intersectional Sociolinguistics*.

Podesva, Robert J. & Patrick Callier. 2015. Voice Quality and Identity. *Annual Review of Applied Linguistics* 35(2015). 173–194.

Pratt, Teresa. 2023. Affect in sociolinguistic style. *Language in Society* 52(1). 1–26.

Pressman, Joel J. 1942. Physiology of the vocal cords in phonation and respiration. *Archives of Otolaryngology* 35. 355–98.

Price, P.J. 1989. Male and female voice source characteristics: inverse filtering results. *Speech Communication* 8(3). 261–277.

Qualtrics. 2005. *Qualtrics*. Version January 2023. Provo, Utah, USA.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Raj, Anoop, Bulbul Gupta, Anindita Chowdhury & Shelly Chadha. 2010. A study of voice changes in various phases of menstrual cycle and in postmenopausal women. *Journal of Voice* 24(3). 363–368.

Ralph in danger. 2014. [Originally from *Family Guy* episode 'The Simpsons Guy', aired September 28th, 2014].

Raymond, Janice. 1979. *The transsexual empire: the making of the she-male*. Boston: Beacon Press.

Rengaswamy, Pradeep, M Gurunath Reddy, K Sreenivasa Rao & Pallab Dasgupta. 2021. Hf0: a hybrid pitch extraction method for multimodal voice. *Circuits, Systems, and Signal Processing* 40(1). 262–275.

Roach, Peter. 2004. British English: received pronunciation. *Journal of the International Phonetic Association* 34(2). 239–245.

Rojas, Sandra, Elaina Kefalianos & Adam Vogel. 2020. How does our voice change as we age? a systematic review and meta-analysis of acoustic and perceptual voice data from healthy adults over 50 years of age. *Journal of Speech, Language, and Hearing Research* 63(2). 533–551.

Rolling Stone. 2008. 100 greatest singers of all time (2008). *Rolling Stone*.

Rowling, J.K. 2022. *My article for the Sunday Times Scotland on why I oppose Gender Recognition Act reform*. https://www.jkrowling.com/opinions/my-article-for-the-sunday-times-scotland-on-why-i-oppose-gender-recognition-act-reform/. Blog post.

Saleem, Mehak. 2020. Sociophonetic stratification of punjabi trans women in lahore. *Pakistan Journal of Language Studies* 4.

San Segundo, Eugenia, Paul Foulkes, Peter French, Philip Harrison, Vincent Hughes & Colleen Kavanagh. 2019. The use of the vocal profile analysis for speaker characterization: methodological proposals. *Journal of the International Phonetic Association* 49(3). 353–380.

Sapienza, Christine M & Jeniffer Dutka. 1996. Glottal airflow characteristics of women's voice production along an aging continuum. *Journal of Speech, Language, and Hearing Research* 39(2). 322–328.

Schaeffler, Felix, Matthias Eichner & Janet Beck. 2019. Towards ordinal classification of voice quality features with acoustic parameters. In *Proceedings of the conference on electronic speech signal processing, TU Dresden, 6-8 march 2019 (ESSV2019)*. 288–295.

Schleef, Erik. 2013. Glottal replacement of /t/ in two British capitals: Effects of word frequency and morphological compositionality. *Language Variation and Change* 25(2). 201–223.

Schmitt, Holger. 2015. Orkney English phonology: observations from interview data. *Scottish Language* 34. 58+.

Segundo, Eugenia San, Paul Foulkes, Peter French, Philip Harrison, Vincent Hughes & Colleen Kavanagh. 2018. Cluster analysis of voice quality ratings: identifying groups of perceptually similar speakers. In *Proceedings of phonetics and phonology in the German-speaking countries(P&P13)*, 173–176. Berlin: AG Elektronisches Publizieren.

Segundo, Eugenia San & Jose A Mompean. 2017. A Simplified Vocal Profile Analysis Protocol for the Assessment of Voice Quality and Speaker Similarity. *Journal of Voice* 31(5). 644.e11–644.e27.

Shue, Yen-Liang, 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), Hong Kong*, 1846–1849.

Signorello, Rosario, Nari Rhee, Bruce Gerratt & Jody Kreiman. 2016. Toward a psychoacoustic model of spectral noise in the voice source. In *10th International Conference on Voice Physiology and Biomechanics(ICVPB), Vina del Mar, Chile*.

Silverstein, Michael. 1976. Shifters, linguistic categories, and cultural description. *Meaning in Anthropology*. 11–55.

Silverstein, Michael. 2003. Indexical order and the dialectics of sociolinguistic life. *Language and Communication* 23(3-4). 193–229.

Simpson, Adrian P. 2012. The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics* 40(3). 477–490.

Sjölander, Kåre. 2004. *The Snack Sound Toolkit*.

Slifka, Janet. 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice* 20(2). 171–186.

Smith, Jennifer, David Adger, Brian Aitken, Caroline Heycock, E Jamieson & Gary Thoms. 2019. *The Scots Syntax Atlas*. https://scotssyntaxatlas.ac.uk.

Smith, Jennifer & Mercedes Durham. 2011. A tipping point in dialect obsolescence? Change across the generations in Lerwick, Shetland. *Journal of Sociolinguistics* 15(2). 197–225.

Smith, Jonathan. 1996. Beyond the divide between cognition and discourse: using interpretative phenomenological analysis in health psychology. *Psychology and Health* (11). 261–271.

Smith, Jonathan A., Paul Flowers & Michael Larkin. 2009. *Interpretive phenomenological analysis: theory, method and research*. SAGE Publications Ltd.

Smith, Jonathan A., Paul Flowers & Michael Larkin. 2022. *Interpretative phenomenological analysis: theory, method and research*. 2nd Core new. Thousand Oaks: SAGE Publications Ltd.

Sonderegger, Morgan. 2021. *Regression Modeling for Linguistic Data*.

Starr, Rebecca L. 2015. Sweet voice: The role of voice quality in a Japanese feminine style. *Language in Society* 44(1). 1–34.

Steele, Ariana. 2022. Intersectionality of social meaning: race, gender and /s/ perception. In Stanford, USA.

Stevens, Kenneth. 1977. Physics of laryngeal behavior and larynx modes. *Phonetica* 34. 264–279.

Stevens, Kenneth Noble & A. S. House. 1961. An acoustical theory of vowel production and some of its implications. *Journal of Speech and Hearing Research* 4. 303–320.

Stone, Sandy. 2013. The empire strikes back: a posttranssexual manifesto. In *The transgender studies reader*, 221–235. Routledge.

Stross, Brian. 2013. Falsetto voice and observational logic: Motivated meanings. *Language in Society* 42(2). 139–162.

Stryker, Susan. 1994. My words to Victor Frankenstein above the village of Chamounix: performing transgender rage. 1. 237–254.

Stuart-Smith, Jane. 1999a. Glasgow: Accent and voice quality. In Paul Foulkes & Gerard Docherty (eds.), *Urban voices: accent studies in the British Isles*, 201–222. Leeds: Arnold.

Stuart-Smith, Jane. 1999b. Voice quality in Glaswegian. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 2553–2556.

Sun, Xuejing. 2002. Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2002, May 13-17 2002, Orlando, Florida, USA*.

Sundkvist, Peter. 2011. "Standard English" as spoken in Shetland's capital. *World Englishes* 30(2). 166–181.

Švec, Jan G. & Svante Granqvist. 2010. Guidelines for Selecting Microphones for Human Voice Production Research. *American Journal of Speech-Language Pathology* 19(356-368).

Sweet, Henry. 1902. *A primer of phonetics*. English. 2nd, rev. Oxford: Clarendon Press.

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*, vol. 1, 438–441.

Szakay, Anita. 2012. Voice quality as a marker of ethnicity in New Zealand: From acoustics to perception. *Journal of Sociolinguistics* 16(3). 382–397.

Szakay, Anita & 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), Glasgow, Scotland*.

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

Talkin, David. 1995. A robust algorithm for pitch tracking. In W. B. Kleijn & K. K. Paliwal (eds.), *Speech coding and synthesis*, 495–518. Philadelphia: Elsevier Science.

Talkin, David. 2015. *REAPER: Robust Epoch And Pitch EstimatoR*.

Thomson, Rachel, Jessica Baker & Julie Arnot. 2018. *Health Care Needs Assessment of Gender Identity Services*.

Tian, Jia & Jianjing Kuang. 2021. The phonetic properties of the non-modal phonation in Shanghainese. *Journal of the International Phonetic Association* 51(2). 202–228.

Titze, Ingo R. 1989. Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America* 85. 1699–1707.

Trudgill, Peter. 1974. *The social differentiation of English in Norwich*. Vol. 13. London & New York: Cambridge University Press.

Tsanas, Athanasios, Matías Zañartu, Max A. Little, Cynthia Fox, Lorraine O. Ramig & Gari D. Clifford. 2014. Robust fundamental frequency estimation in sustained vowels: Detailed algorithmic comparisons and information fusion with adaptive Kalman filtering. *The Journal of the Acoustical Society of America* 135(5). 2885–2901.

Turk, Alice, Satsuki Nakai & Mariko Sugahara. 2006. Acoustic segment durations in prosodic research: a practical guide. In *Methods in empirical prosody research*, 1–28. Mouton de Gruyter.

Venables, William & Brian Ripley. 2002. *Modern applied statistics with s*. Fourth. New York: Springer.

Vicenik, Chad, Spencer Lin, Patricia Keating & Yen-Liang Shue. 2022. *Online documentation for VoiceSauce*.

Vieira, M. N. 1997. *Automated measures of dysphonias and the phonator effects of asymmetries in the posterior larynx*. University of Edinburgh dissertation.

Webb, A L, P N Carding, I J Deary, K MacKenzie, N Steen & J A Wilson. 2004. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol* 261(8). 429–434.

Wells, J C. 1982. *Accents of English: The British Isles*. Cambridge: Cambridge University Press.

White, Hannah, Joshua Penney, Andy Gibson, Anita Szakay & Felicity Cox. 2022. Evaluating automatic creaky voice detection methods. *The Journal of the Acoustical Society of America* 152(3). 1476–1486.

Wickham, Hadley. 2016. *Ggplot2: elegant graphics for data analysis*. Springer-Verlag New York.

Wileman, Rory Bruce. 2018. *A sociophonetic investigation of ethnolinguistic differences in voice quality among young, South African English speakers*. University of Cape Town dissertation.

Wiley, Matt & Joshua Wiley. 2019. *Advanced R statistical programming and data models: analysis, machine learning, and visualization*. English. Berkeley, CA: Apress.

Wilson, D K. 1987. *Voice problems of children*. 3rd. Baltimore: Williams & Wilkins.

Wilson, Ian & Bryan Gick. 2014. Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research* 57(2). 361–373.

van der Woerd, Benjamin, Min Wu, Vijay Parsa, Philip C Doyle & Kevin Fung. 2020. Evaluation of acoustic analyses of voice in nonoptimized conditions. *Journal of Speech Language and Hearing Research* 63. 3991–3999.

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

Yanushevskaya, Irena, Christer Gobl & Ailbhe Ní Chasaide. 2018. Cross-language differences in how voice quality and f0 contours map to affect. *The Journal of the Acoustical Society of America* 144(5). 2730–2750.

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

Yumoto, E., W. J. Gould & T. Baer. 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. English. *The Journal of the Acoustical Society of America* 71. 1544–1550.

Zanghellini, Aleardo. 2020. Philosophical problems with the gender-critical feminist argument against trans inclusion. *SAGE Open* 10(2).

Zemlin, W. R. 1964. *Speech and hearing science*. Champaign, Illinois: Stipes.

Zhang, Zhaoyan. 2016a. Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America* 139.

Zhang, Zhaoyan. 2016b. Mechanics of human voice production and control. *The Journal of the Acoustical Society of America* 140(4). 2614–2635.

Zimman, Lal. 2012. *Voices in transition: testosterone, transmasculinity, and the gendered voice among female-to-male transgender people*. University of Colorado dissertation.

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

Zimman, Lal. 2017a. Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/. *Language in Society* 46(3). 339–370.

Zimman, Lal. 2017b. Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants. *Linguistics* 55(5). 993–1019.

Zimman, Lal. 2021. Gender diversity and the voice. In Judith Baxter & Jo Angouri (eds.), *The Routledge handbook of language, gender, and sexuality*. New York: Routledge.

Appendix A

Additional material relating to Part II

A.1 Additional contingency tables for PPA results

Contingency tables were generated using the janitor package (Firke 2020) and exported using xtable (Dahl et al. 2019).

A.2 Overall distributions of scalar degrees

Table A.1: Distribution of scalar degrees of whispery creaky voice, exclusively whispery voice and exclusively breathy voice across all voiced stretches

Voice quality	1	2	3	4	5	Total
Whispery creaky						
Creaky	81% (595)	6% (44)	2% (18)	2% (18)	8% (62)	100% (737)
Whispery	26% (192)	41% (300)	21% (153)	9% (67)	3% (25)	100% (737)
Breathy	28% (104)	22% (82)	32% (120)	12% (45)	5% (19)	100% (370)
Whispery	15% (96)	30% (186)	38% (237)	12% (76)	4% (25)	100% (620)
Creaky	83% (205)	7% (18)	4% (10)	1% (3)	4% (10)	100% (246)

A.2.1 Social factors

Table A.2: Distribution of all voice qualities by area

Voice quality	Glasgow	Lothian	Shetland
breathy	12% (92)	32% (214)	9% (64)
creaky	4% (33)	21% (138)	10% (75)
creaky_breathy	1% (10)	1% (7)	0% (1)
creaky_falsetto	0% (1)	0% (0)	0% (0)
creaky_harsh	0% (0)	0% (1)	0% (0)
harsh	0% (1)	1% (5)	0% (0)
modal	0% (2)	12% (78)	6% (46)
whisper	0% (1)	0% (1)	0% (1)
whispery	36% (276)	14% (92)	35% (252)
whispery_creaky	45% (344)	17% (117)	38% (276)
whispery_creaky_falsetto	0% (1)	0% (0)	0% (0)
whispery_creaky_harsh	0% (0)	0% (3)	0% (0)
whispery_falsetto	0% (1)	0% (0)	0% (0)
whispery_harsh	1% (5)	3% (17)	2% (15)
Total	100% (767)	100% (673)	100% (730)

Table A.3: Distribution of all voice qualities by gender

Voice quality	F	M
breathy	3% (37)	31% (333)
creaky	16% (174)	7% (72)
creaky_breathy	0% (0)	2% (18)
creaky_falsetto	0% (1)	0% (0)
creaky_harsh	0% (0)	0% (1)
harsh	0% (2)	0% (4)
modal	4% (49)	7% (77)
whisper	0% (1)	0% (2)
whispery	33% (356)	24% (264)
whispery_creaky	43% (464)	25% (273)
whispery_creaky_falsetto	0% (0)	0% (1)
whispery_creaky_harsh	0% (0)	0% (3)
whispery_falsetto	0% (1)	0% (0)
whispery_harsh	1% (6)	3% (31)
Total	100% (1091)	100% (1079)

Table A.4: Distribution of all voice qualities by age group

Voice quality	O	Y
breathy	16% (178)	18% (192)
creaky	12% (131)	11% (115)
creaky_breathy	1% (7)	1% (11)
creaky_falsetto	0% (0)	0% (1)
creaky_harsh	0% (1)	0% (0)
harsh	0% (4)	0% (2)
modal	3% (38)	8% (88)
whisper	0% (1)	0% (2)
whispery	32% (358)	25% (262)
whispery_creaky	33% (374)	35% (363)
whispery_creaky_falsetto	0% (0)	0% (1)
whispery_creaky_harsh	0% (3)	0% (0)
whispery_falsetto	0% (0)	0% (1)
whispery_harsh	3% (36)	0% (1)
Total	100% (1131)	100% (1039)

Table A.5: Distribution of all voice qualities by glottal context

Voice quality	Not glottal % (n)	Glottal % (n)
Breathy	18 (333)	11 (37)
Creaky	10 (189)	17 (57)
Creaky breathy	0 (6)	4 (12)
Creaky falsetto	0 (1)	0 (0)
Creaky harsh	0 (1)	0 (0)
Harsh	0 (5)	0 (1)
Modal	6 (116)	3 (10)
Whisper	0 (3)	0 (0)
Whispery	31 (561)	17 (59)
Whispery creaky	32 (587)	44 (150)
Whispery creaky falsetto	0 (0)	0 (1)
Whispery creaky harsh	0 (2)	0 (1)
Whispery falsetto	0 (1)	0 (0)
Whispery harsh	1 (26)	3 (11)
Total	100 (1831)	100 (339)

Table A.6: Distribution of all voice qualities by phrase position (initial or non-initial)

Voice quality	not initial	initial
breathy	18% (274)	15% (96)
creaky	10% (156)	14% (90)
creaky_breathy	1% (8)	2% (10)
creaky_falsetto	0% (1)	0% (0)
creaky_harsh	0% (0)	0% (1)
harsh	0% (5)	0% (1)
modal	6% (93)	5% (33)
whisper	0% (1)	0% (2)
whispery	29% (455)	27% (165)
whispery_creaky	34% (528)	34% (209)
whispery_creaky_falsetto	0% (0)	0% (1)
whispery_creaky_harsh	0% (2)	0% (1)
whispery_falsetto	0% (1)	0% (0)
whispery_harsh	2% (25)	2% (12)
Total	100% (1549)	100% (621)

Table A.7: Distribution of all voice qualities by phrase position (final or non-final)

Voice quality	not final	final
breathy	18% (295)	15% (75)
creaky	10% (168)	16% (78)
creaky_breathy	1% (16)	0% (2)
creaky_falsetto	0% (1)	0% (0)
creaky_harsh	0% (1)	0% (0)
harsh	0% (4)	0% (2)
modal	5% (91)	7% (35)
whisper	0% (2)	0% (1)
whispery	29% (492)	26% (128)
whispery_creaky	34% (570)	33% (167)
whispery_creaky_falsetto	0% (0)	0% (1)
whispery_creaky_harsh	0% (3)	0% (0)
whispery_falsetto	0% (1)	0% (0)
whispery_harsh	2% (26)	2% (11)
Total	100% (1670)	100% (500)

Table A.8: Contingency table showing degree of whispery voice by social factor in voiced stretches coded as exclusively whispery

Factor	1	2	3	4	5	Total
Area						
Glasgow	7% (18)	25% (70)	47% (129)	16% (44)	5% (15)	100% (276)
Lothian	20% (18)	30% (28)	46% (42)	3% (3)	1% (1)	100% (92)
Shetland	24% (60)	35% (88)	26% (66)	12% (29)	4% (9)	100% (252)
Gender						
F	19% (69)	32% (113)	39% (139)	7% (25)	3% (10)	100% (356)
M	10% (27)	28% (73)	37% (98)	19% (51)	6% (15)	100% (264)
Age group						
O	15% (54)	29% (103)	35% (127)	15% (55)	5% (19)	100% (358)
Y	16% (42)	32% (83)	42% (110)	8% (21)	2% (6)	100% (262)

Table A.9: Contingency table showing degree of whispery voice by linguistic factor in voiced stretches coded as exclusively whispery

Factor	1	2	3	4	5	Total
Glottal context						
Not glottal	15% (85)	29% (164)	40% (223)	12% (68)	4% (21)	100% (561)
Glottal	19% (11)	37% (22)	24% (14)	14% (8)	7% (4)	100% (59)
Phrase initial						
Not initial	16% (74)	27% (123)	38% (175)	13% (61)	5% (22)	100% (455)
Initial	13% (22)	38% (63)	38% (62)	9% (15)	2% (3)	100% (165)
Phrase final						
Not final	16% (77)	30% (146)	38% (185)	13% (63)	4% (21)	100% (492)
Final	15% (19)	31% (40)	41% (52)	10% (13)	3% (4)	100% (128)

Table A.10: Contingency table showing degree of whispery voice and creaky voice by social factor in voiced stretches coded as whispery creaky voice

Factor	Voice quality	1	2	3	4	5	Tota
Gender							
F	Creaky	83% (387)	6% (28)	2% (11)	2% (7)	7% (31)	100%
	Whispery	29% (135)	43% (198)	20% (91)	7% (34)	1% (6)	100%
M	Creaky	76% (208)	6% (16)	3% (7)	4% (11)	11% (31)	100%
	Whispery	21% (57)	37% (102)	23% (62)	12% (33)	7% (19)	100%
Area							
Glasgow	Creaky	84% (288)	4% (15)	2% (6)	3% (12)	7% (23)	100%
	Whispery	24% (82)	46% (158)	19% (66)	8% (29)	3% (9)	100%
Lothian	Creaky	89% (104)	2% (2)	3% (4)	1% (1)	5% (6)	100%
	Whispery	37% (43)	43% (50)	15% (17)	4% (5)	2% (2)	100%
Shetland	Creaky	74% (203)	10% (27)	3% (8)	2% (5)	12% (33)	100%
	Whispery	24% (67)	33% (92)	25% (70)	12% (33)	5% (14)	100%
Age group							
O	Creaky	85% (318)	5% (19)	2% (9)	2% (6)	6% (22)	100%
	Whispery	24% (91)	45% (167)	17% (63)	10% (37)	4% (16)	100%
Y	Creaky	76% (277)	7% (25)	2% (9)	3% (12)	11% (40)	100%
	Whispery	28% (101)	37% (133)	25% (90)	8% (30)	2% (9)	100%

Table A.11: Contingency table showing degree of whispery voice and creaky voice by linguistic factor in voiced stretches coded as whispery creaky voice

Factor	Voice quality	1	2	3	4	5
Glottal context						
Not glottal	creaky	85% (500)	5% (27)	2% (13)	1% (8)	7% (39)
Not glottal	whispery	25% (148)	42% (247)	21% (124)	8% (48)	3% (20)
Glottal	creaky	63% (95)	11% (17)	3% (5)	7% (10)	15% (23)
Glottal	whispery	29% (44)	35% (53)	19% (29)	13% (19)	3% (5)
Phrase initial						
Not initial	creaky	82% (431)	6% (30)	2% (13)	2% (13)	8% (41)
Not initial	whispery	26% (135)	41% (216)	22% (116)	9% (45)	3% (16)
Initial	creaky	78% (164)	7% (14)	2% (5)	2% (5)	10% (21)
Initial	whispery	27% (57)	40% (84)	18% (37)	11% (22)	4% (9)
Phrase final						
Not final	creaky	82% (470)	6% (32)	2% (13)	2% (14)	7% (41)
Not final	whispery	27% (152)	40% (229)	21% (117)	9% (52)	4% (20)
final	creaky	75% (125)	7% (12)	3% (5)	2% (4)	13% (21)
final	whispery	24% (40)	43% (71)	22% (36)	9% (15)	3% (5)

A.2.2 Linguistic factors

A.3 Exclusively whispery voice

A.4 Whispery creaky voice

A.5 Exclusively breathy voice

Table A.12: Contingency table showing degree of breathy voice by social factor in voiced stretches coded as exclusively breathy voice

	1	2	3	4	5	Total
Area						
Glasgow	7% (6)	29% (27)	40% (37)	18% (17)	5% (5)	100% (92)
Lothian	42% (89)	21% (44)	29% (62)	7% (15)	2% (4)	100% (214)
Shetland	14% (9)	17% (11)	33% (21)	20% (13)	16% (10)	100% (64)
Gender						
F	62% (23)	24% (9)	11% (4)	3% (1)	0% (0)	100% (37)
M	24% (81)	22% (73)	35% (116)	13% (44)	6% (19)	100% (333)
Age group						
O	28% (49)	16% (29)	35% (62)	15% (26)	7% (12)	100% (178)
Y	29% (55)	28% (53)	30% (58)	10% (19)	4% (7)	100% (192)

Table A.13: Contingency table showing degree of breathy voice by linguistic factor in voiced stretches coded as exclusively breathy voice

	1	2	3	4	5	Total
Glottal context						
Not glottal	28% (94)	22% (74)	32% (107)	12% (39)	6% (19)	100% (333)
Glottal	27% (10)	22% (8)	35% (13)	16% (6)	0% (0)	100% (37)
Phrase initial						
not initial	28% (77)	23% (63)	31% (86)	12% (33)	5% (15)	100% (274)
initial	28% (27)	20% (19)	35% (34)	12% (12)	4% (4)	100% (96)
Phrase final						
not final	27% (79)	21% (63)	33% (97)	14% (41)	5% (15)	100% (295)
final	33% (25)	25% (19)	31% (23)	5% (4)	5% (4)	100% (75)

A.6 Exclusively creaky voice

Table A.14: Contingency table showing degree of creaky voice by social factor in voiced stretches coded as exclusively creaky voice

	1	2	3	4	5	Total
Area						
Glasgow	85% (28)	3% (1)	3% (1)	0% (0)	9% (3)	100% (33)
Lothian	83% (115)	4% (6)	7% (9)	1% (2)	4% (6)	100% (138)
Shetland	83% (62)	15% (11)	0% (0)	1% (1)	1% (1)	100% (75)
Gender						
F	83% (145)	5% (8)	6% (10)	1% (2)	5% (9)	100% (174)
M	83% (60)	14% (10)	0% (0)	1% (1)	1% (1)	100% (72)
Age group						
O	84% (110)	5% (7)	6% (8)	2% (2)	3% (4)	100% (131)
Y	83% (95)	10% (11)	2% (2)	1% (1)	5% (6)	100% (115)

Table A.15: Contingency table showing degree of creaky voice by linguistic factor in voiced stretches coded as exclusively creaky voice

	1	2	3	4	5	Total
Glottal context						
Not glottal	89% (168)	5% (9)	3% (5)	1% (2)	3% (5)	100% (189)
Glottal	65% (37)	16% (9)	9% (5)	2% (1)	9% (5)	100% (57)
Phrase initial						
Not initial	83% (130)	8% (12)	4% (6)	2% (3)	3% (5)	100% (156)
Initial	83% (75)	7% (6)	4% (4)	0% (0)	6% (5)	100% (90)
Phrase final						
Not final	83% (139)	8% (13)	4% (7)	2% (3)	4% (6)	100% (168)
Final	85% (66)	6% (5)	4% (3)	0% (0)	5% (4)	100% (78)

Appendix B

Interpreting multinomial logit models: a tutorial

As multinomial logit models are not commonly used in linguistics, a tutorial on how to interpret the results of them can be found in the supplementary materials available through OSF here.

B.1 Within-category ordered logit for breathy voice

Table B.1: Summary of ordered logit model predicting the degree of breathy voice as a function of H1*-H2*, H2*-H4* and HNR05.

	Dependent variable:
	breathy
H1H2c_scaled	0.447
	(0.100)
	t = 4.488***
H2H4c_scaled	0.465
	(0.108)
	t = 4.295***
HNR05_scaled	−0.172
	(0.101)
	t = −1.704
Log Likelihood	-428.7901
Observations	335
Note:	*p<0.05; **p<0.01; ***p<0.001

Table B.1 presents the results of the ordered logit model predicting the probability of a breathy voiced stretch being rated with increasing degrees of breathy voice, given H1*-H2*, H2*-H4* and HNR05. All independent variables have been mean-centred and scaled.

B.1.0.1 H1*-H2*

Table B.2: Predicted probabilities for values of H1*-H2* ranging between -2 and +2 standard deviations from the mean of H1*-H2*, with values for all other predictors held constant at their means.

Scaled H1*-H2*	Raw H1*-H2*	1	2	3	4
-2	-2.99	0.48	0.25	0.21	0.06
-1	0.92	0.37	0.26	0.28	0.08
0	4.82	0.28	0.25	0.35	0.13
1	8.73	0.20	0.22	0.40	0.18
2	12.64	0.14	0.18	0.43	0.26

Figure B.1 shows the effect of H1*-H2* on the predicted probability of a breathy voiced stretch being rated as a particular scalar degree. Predicted probabilities for a range of values for H1*-H2* are shown in B.2.

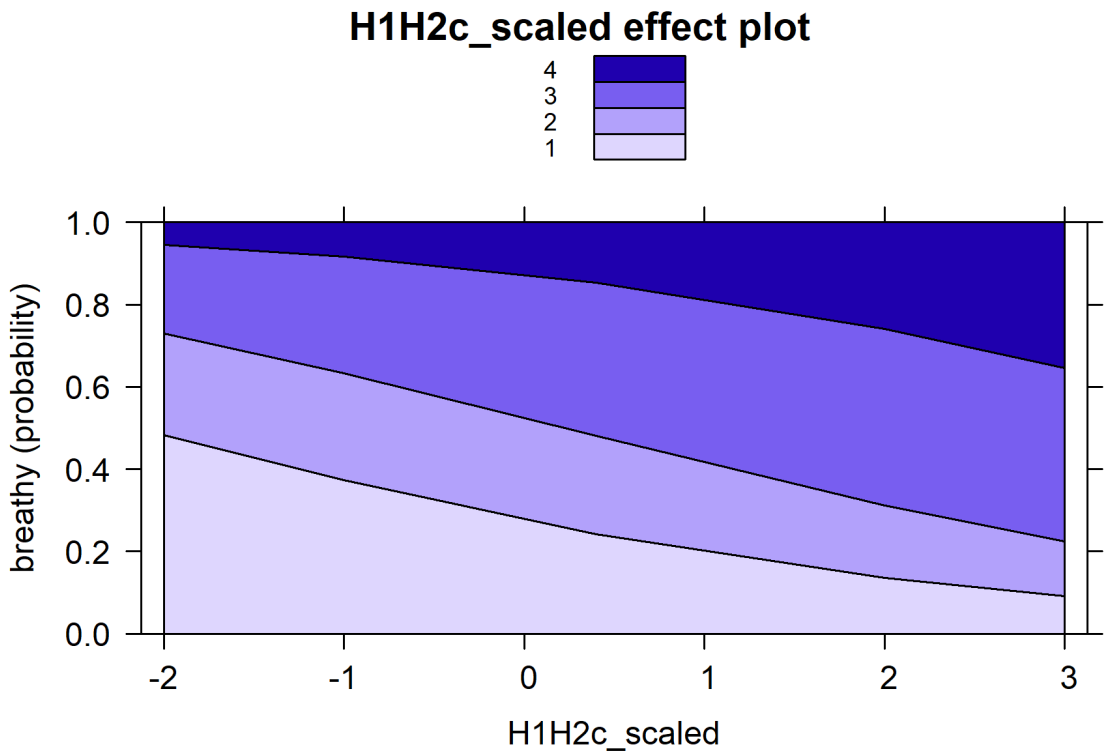


Figure B.1: Effect of H1*-H2* on the predicted probability of a breathy voiced stretch being rated as a particular scalar degree (all other independent variables held constant at their means)

For a 1 SD increase in H1*-H2*, the odds of a breathy voiced stretch being rated as increasingly breathy change by 1.563 while the other variables in the model are held constant, representing a significant increase.

B.1.0.2 H2*-H4*

Table B.3: Predicted probabilities for values of H2*-H4* ranging between -2 and +2 standard deviations from the mean of H2*-H4*, with values for all other predictors held constant at their means.

Scaled H2*-H4*	Raw H2*-H4*	1	2	3	4
-2	-2.42	0.49	0.25	0.21	0.05
-1	1.48	0.38	0.26	0.28	0.08
0	5.38	0.28	0.25	0.35	0.13
1	9.29	0.19	0.22	0.40	0.19
2	13.19	0.13	0.17	0.43	0.27

Figure B.2 shows the effect of H2*-H4* on the predicted probability of a breathy voiced stretch being rated as a particular scalar degree. Predicted probabilities for a range of values for H2*-H4* are shown in B.3.

For a 1 SD increase in H2*-H4*, the odds of a breathy voiced stretch being rated as increasingly breathy increase by 1.592, while the other variables in the model are

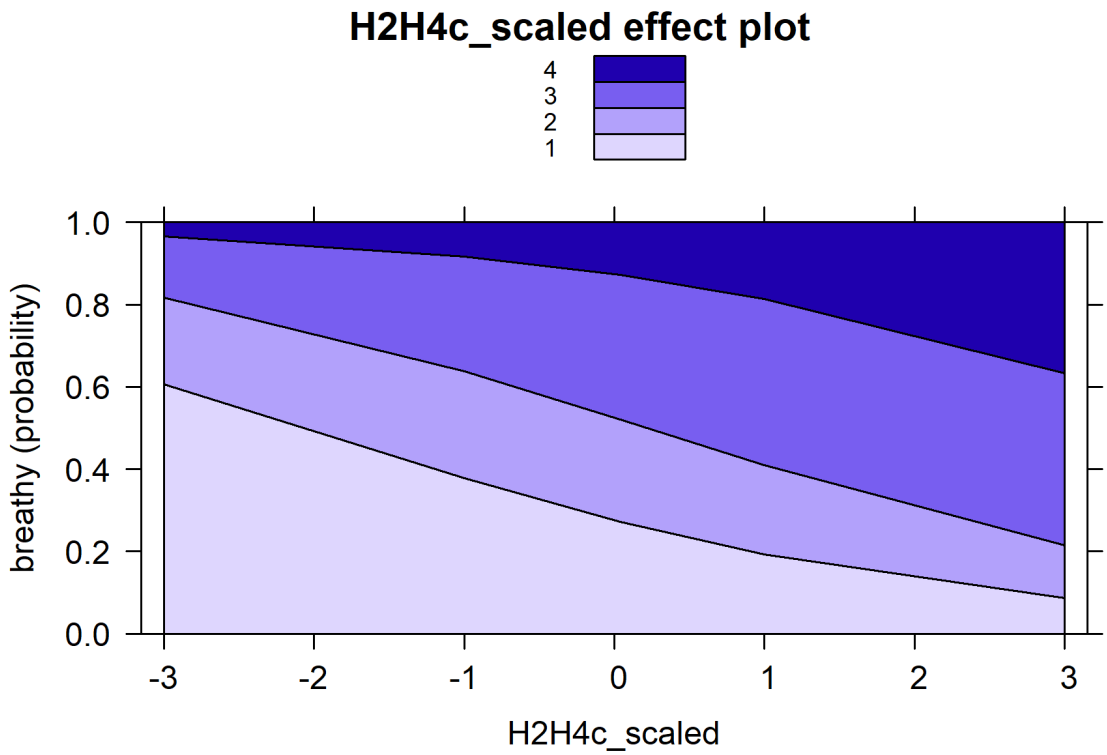


Figure B.2: Effect of H2*-H4* on the predicted probability of a breathy voiced stretch being rated as a particular scalar degree (all other independent variables held constant at their means)

held constant.

B.1.0.3 HNR05

There was no significant effect of HNR05 on the odds of a breathy voiced stretch being rated as increasingly breathy.

B.2 Within-category ordered logit for whispery voice

Table B.4 presents the results of the ordered logit model predicting the probability of a whispery voiced stretch being rated with increasing degrees of breathy voice, given HNR15. All independent variables have been mean-centred and scaled. Predicted probabilities for a range of values for HNR15 are presented in Table B.5.

Figure B.3 shows the effect of HNR15 on the predicted probability of a whispery voiced stretch being rated as a particular scalar degree. For a 1 SD increase in HNR15, odds of a breathy voiced stretch being rated as increasingly breathy change by 0.678, representing a significant decrease.

Table B.4: Summary of ordered logit model predicting the degree of whispery voice as a function of HNR15.

	Dependent variable:
	whispery
HNR15_scaled	−0.388
	(0.079)
	t = −4.915***
Log Likelihood	-701.3547
Observations	549
Note:	*p<0.05; **p<0.01; ***p<0.001

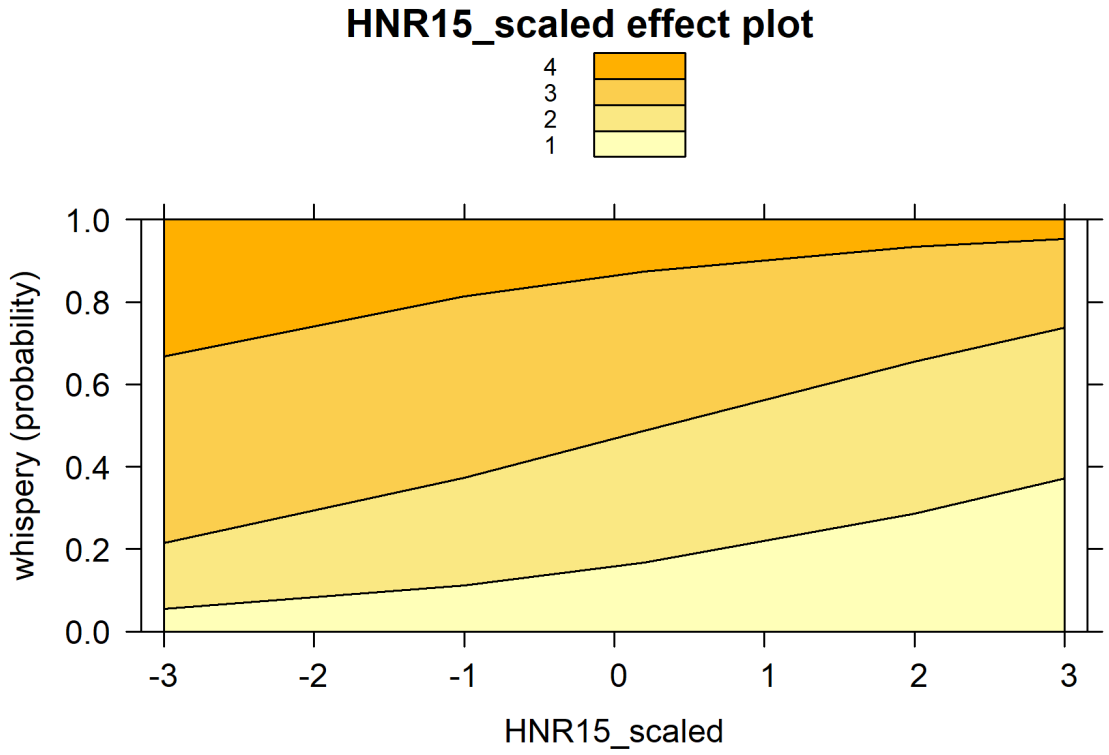


Figure B.3: Effect of HNR15 on the predicted probability of a whispery voiced stretch being rated as a particular scalar degree (all other independent variables held constant at their means)

Table B.5: Probability of breathy voiced stretches being rated as each degree of whispery, as predicted by the multinomial logit model, for a range of values for HNR15.

Scaled HNR15	HNR15	1	2	3	4
-2	23.24	0.08	0.21	0.46	0.25
-1	31.08	0.11	0.26	0.44	0.19
0	38.92	0.16	0.31	0.40	0.13
1	46.76	0.21	0.35	0.34	0.10
2	54.59	0.29	0.37	0.28	0.07

Appendix C

Additional materials from Part III

C.1 Ethical approval

Ethical approval removed due to confidentiality issues.

C.2 IPA interview schedule

Research question: How do these participants understand their experiences with changes in their voices as transgender individuals?

Topics covered:

- The aspects of the voice that participants understand to be involved in changes in their voice
- How participants understand the relationship between their experiences of voice change and their interpersonal relationships
- The physical and embodied experiences of voice changes
- How participants understand the relationship between their experiences of voice change and their identities

Describing the experience of voice change

Can you tell me a bit about what your experience with your voice has been like?

> *Prompts:*
>
> Do you think that your voice has changed over time?
>
> Could you describe how you think your voice might have changed?
>
> What has the time frame of this been – is it more gradual or sudden?
>
> How intentional does it feel when your voice changes?
>
> What happens when it changes?
>
> What does it feel like?

Have you ever done anything to try to change the way your voice sounds?

If yes:

Could you tell me a bit about the things that you've done to change the sound of your voice?

> *Prompts:*
>
> Some people use voice training apps, watch YouTube videos, take testosterone or go to voice therapy – Have you done anything like that? Have you tried anything else?
>
> Can you describe the what it's like when you…?
>
> How helpful has x been? How has x helped?
>
> What does 'progress' look like?
>
> What was it like when you started?
>
> How do you feel when you're on your way to an appointment/doing the exercises?
>
> What does it feel like physically?
>
> How long have you been doing x for?
>
> Could you tell me about your decision to start/not start doing x?

Are there any other things that you would like to try?

It sounds like you would like to do x but haven't been able to – can you tell me a bit more about that

If no:

Could you tell me a bit more why you haven't?

Your voice and interpersonal relationships

As you know, I recorded you speaking to two different people. Do you think that your voice changed at all between those two conversations, and if so, how?

Prompts:

Could you tell me about why you think your voice changed between those two conversations?

Could you tell me how you think your voice affects the way that the people you were speaking to see you?

Can you tell me about a time/another time that you think your voice changed when you spoke to different people?

How much control do you feel like you have over this?

Could you tell me a bit about what it feels like when your voice changes in this way?

What happens when you're speaking to someone you're more/less comfortable with?

Can you tell me about how the context that you're speaking to someone in affects your voice? By this I mean things like the place that you're in, whether you're calling someone or speaking to them in person, if you're in a group of people or one-on-one with someone, and things like that.

Prompts:

Can you tell me what it's like when you speak to someone over the phone/on voice chat compared to if they can see your face?

Can you tell me a bit about what it's like when you're doing public speaking?

What's it like at work compared to with friends?

What's it like in public compared to at home?

What about if you're meeting someone new?

Your voice and your identity

How well do you think your voice represents who you are?

Prompts:

How well do you think it fits you?

How well do you feel you can express your feelings with your voice?

What do you think you would need for it to fit you better?

C.3 Conversation starters for recorded conversations

How do you know each other?

What's the most fun you've had together?

If you could go on holiday anywhere together, where would it be and why?

As a child, did you ever get blamed for something you didn't do? What happened?

What's your weirdest recurring dream?

What was your first job?

What was your worst job?

Are you more productive in the morning or the evening?

Have you ever stayed up all night for work or studying?

What subject should be taught in school but isn't?

Did you have any really weird teachers in school?

Where exactly are you from? Was it a good place to grow up?

Have you ever lived anywhere else? How does it compare to where you live now?

Do you think other people can tell where you're from when they hear you speak?

If you could go on holiday anywhere, where would it be and why?

Have you ever had a really big argument? How did you figure it out?

Figure out something you have in common that you didn't know about before

As a child, did you ever get blamed for something you didn't do? What happened?

What's your weirdest recurring dream?

What was your first job?

What was your worst job?

Are you more productive in the morning or the evening?

Have you ever stayed up all night for work or studying?

What subject should be taught in school but isn't?

Did you have any really weird teachers in school?

Where exactly are you from? Was it a good place to grow up?

Have you ever lived anywhere else? How does it compare to where you live now?

Do you think other people can tell where you're from when they hear you speak?

C.4 Participant Information Sheet and Consent Form

University of Glasgow

College of Arts
Research Ethics

Participant Information FAQ: Plain Language Statement

Trans individuals' experiences with their voices

Researcher: Joe Pearce

Supervisor: Prof. Jane Stuart-Smith, Dr. Clara Cohen, Dr. Felix Schaeffler

Invitation

My name is Joe Pearce. I'm a PhD student at the University of Glasgow interested in trans people's experiences with their voices: Things like if and how someone's voice changes when they speak to different people, or over time, and what these experiences are like for them. I'm interested in this in part because I'm trans myself and I want to hear about the experiences of other trans people.

You are being invited to take part in this research. 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.

What is the purpose of the study?

This research is looking at how trans people understand their experiences with their voices and how they use their voices to speak to different people in their lives.

The research is being conducted as part of my PhD which looks at gender and voices in Scotland more generally. The results will be written up for conference presentations, published in academic papers and used in teaching materials, training materials, and short reports aimed at non-academic audiences. I also plan to communicate the findings to:

- Speech and language therapists, to help them understand the experiences of trans people that they work with
- Organisations campaigning for access to trans healthcare, such as the Scottish Trans Alliance, to provide them with accounts of the impact of access and lack of access to healthcare on trans people's lives
- The National Gender Identity Clinical Network for Scotland (NGICNS), if evidence of shortcomings of gender identity clinics is found, to suggest improvements to the service
- Trans people who are interested in hearing about other people's experiences with their voices

Why have I been chosen?

I'm looking for people from Scotland over the age of 18 who identify as transgender. You were invited to take part because you took part in an earlier part of this research and were invited back to take part in an in-depth interview.

Do I have to take part?

It is up to you to decide whether or not to take part. If you decide to take part, you are still free to withdraw at any time and without giving a reason.

What will happen to me if I take part?

University of Glasgow

College of Arts
Research Ethics

You'll be taking part in an in-depth interview where I'll ask you about topics like how you think your voice might have changed over time, what you think happens to your voice when you speak to different people and what your experiences with your voice have been. I'm calling it an 'interview', but it's not like a job interview – I'm interested in understanding your thoughts on the topic, not judging and evaluating you.
The aim of it is for me to understand your perspective without assuming anything about your experience, so I might ask some questions that sound strange or quite personal. You don't have to answer any questions that you'd like to miss, and you don't have to give a reason for not wanting to answer. This interview may take up to two hours and you will receive £40 for your time.

What if I want to withdraw from the study?

You're free to withdraw from the study at any time while data is being collected, or in the one month following conclusion of the final interview. You don't have to give a reason.

If you withdraw from the study, you will still receive compensation for taking part up until the point where you withdrew. One month after the conclusion of the final interview, you will no longer be able to withdraw from the study, as research outputs may begin to be published after this point.

Will my taking part in this study be kept confidential?

Your participation in this study will be kept confidential. This means your identifying personal data (i.e. your name and contact information) won't be shared outside the research team. You can find more information about how your data will be used and stored in the Privacy Notice here.

What will happen to the project data?

All personal data and research data will be stored securely on a password-protected network drive.

Transcription and anonymization

Any information you disclose in the interview that could potentially identify you will be redacted from the audio recordings in the first month after the recording takes place – any names of people or places will be replaced with silence.

The interviews will then be transcribed using Microsoft's Word for Web Transcribe service. When I do this, the content of the interview will be sent to Microsoft to be transcribed using a computer program – but Microsoft won't store the recording of your interview. The recording that gets transcribed by Word for Web Transcribe will be anonymous and no one will hear it when this is happening.

How I'll analyse the interviews

I'll read over the transcripts of the interviews and listen back to the audio recordings and make notes about what you said, in an attempt to understand your experiences with your voice and how you make sense of them. I'll compare what you said to what other people said in their interviews, and look for similarities and differences between your experiences.

University of Glasgow

College of Arts
Research Ethics

Storing and sharing data after the end of the project

After the project is complete, audio recordings of the interviews will be deleted. However, the consent forms and anonymised transcripts of the interviews will be stored for at least 10 years. If you agree, the transcripts of the interviews resulting from the project will be deposited with the UK Data Archive, the lead organisation of UK Data Service, a trusted digital repository funded by the Economic and Social Research Council (ESRC) to provide access to social science data so that other researchers can use it for research and learning.

The anonymised transcripts of the interviews will be made available for future reuses as **safeguarded permission only data.** This means that researchers will only be able to access the transcripts if they are registered with the UK Data Service. The End User Licence signed at the point of registration ensures that data will be used ethically and responsibly, and researchers agree to. Additionally, due to sensitivities in the data this will only be made available subject to gaining approval from me.

How can I find out about the results of the study?

You can choose whether or not you'd like to be notified when outputs of the study are released. If you do you choose to be notified, your email address will be added to a mailing list so that you can be contacted when a publication is released.

Who is organising and funding the research?

This research is organized by Joe Pearce, and funded by the Economic and Social Research Council (ESRC) through the Scottish Graduate School for Social Science (SGSSS).

Who has reviewed the study?

Materials relating to the study have been reviewed and approved by members of the College of Arts Research Ethics panel.

Application reference number:	**Date of approval letter**:
100220002	15 November 2023

Who can I contact if I have any concerns or complaints about the research project?

You can contact the researcher or their supervisor in the first instance if you have any concerns. If you are not comfortable doing this, or if you have tried but don't get a response or if the person in question appears to have left the University, you can contact the College of Arts Ethics Officer (email: arts-ethics@glasgow.ac.uk).

Contact for further Information

Researcher's name and email:

Joe Pearce

j.pearce.1@research.gla.ac.uk

Supervisor's name and email:

Prof. Jane Stuart-Smith

jane.stuart-smith@glasgow.ac.uk

Department address:

College of Arts
Research Ethics

12 University Gardens G12 8QQ Glasgow
If you have any concerns regarding the conduct of this research project, you can contact the College of Arts Ethics Officer (email: arts-ethics@glasgow.ac.uk).

University of Glasgow

College of Arts
Research Ethics

CONSENT TO PARTICIPATE
AGREEMENT TO THE USE OF DATA

Please tick the appropriate boxes	**Yes**	**No**
1. Taking part in the study		
I understand that Joe Pearce is collecting data in the form of audio-recorded interviews for use in an academic research project at the University of Glasgow.	☐	☐
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	☐	☐
I have read and understood the Participant Information Sheet, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	☐	☐
I have read and understood the Privacy Notice, or it has been read to me. I have been able to ask questions about the use and storage of my data and my questions have been answered to my satisfaction.	☐	☐
2. Use of the information in the study		
I consent to the processing of my personal data for the purposes explained to me in the Information Sheet. I understand that such information will be handled in accordance with the terms of the UK General Data Protection Regulation as described in the Privacy Notice.	☐	☐
I understand that I will be referred to by a pseudonym any outputs that arise from the study. Any names or other identifying information mentioned during the conversation will be redacted or replaced with a generic pseudonym	☐	☐
I understand that research data that this interview generates will be written up in academic publications as well as short reports aimed at healthcare practitioners, campaigners and policy-makers working in trans healthcare, as well as other members of the trans community.	☐	☐
I agree to the interview being transcribed using Microsoft's Word for Web Transcribe function	☐	☐
I agree that the words I say in the interview can be quoted and summarised in research outputs	☐	☐

University of Glasgow

College of Arts
Research Ethics

Yes	No
☐	☐

3. Future use and reuse of the information by others

I give permission for the anonymised **transcripts of the interview** to be deposited with the UK Data Archive. I understand that this data will be **safeguarded permission only**, so will only be available to registered researchers who agree to certain conditions and who are approved by Joe Pearce or someone nominated by them.

_____	_____	_____
Name of participant [IN CAPITALS]	Signature	Date
_____	_____	_____
Name of researcher [IN CAPITALS]	Signature	Date

Researcher's name and email:	Joe Pearce j.pearce.1@research.gla.ac.uk
Supervisor's name and email:	Prof. Jane Stuart-Smith jane.stuart-smith@glasgow.ac.uk
Department address:	12 University Gardens G12 8QQ Glasgow

C.5 Additional statistical models for Carrie's voice for measures with no significant effect for interlocutor

C.6 Creak results

After excluding coding errors, a total of 4336 stretches were coded in Carrie's speech.

Table C.1: Summary of phonation types coded in Carrie's speech

Phonation type	n	% phonation type
Baseline voicing	3232	74.5%
Creak	1066	24.6%
Whisper	35	0.8%
Harsh	3	0.1%
Total	4336	100.0%

Carrie made use of creak, whisper, and forms of harsh voice. Of these, creak was the most common phonation type used, accounting for 74.5% of voiced stretches. Whisper and harsh voice were used comparatively infrequently, both in less than 1% of cases, and so these are not included in the present quantitative analysis. Instead, I will return to these cases in a qualitative acoustic analysis.

After whisper and harsh voice were excluded, this left 4298 stretches for analysis, consisting of 493.2 seconds of voicing. Of this, 69.6 seconds (14.1%) was creaky, and 423.7 seconds (85.9%) was not.

Table C.2: Percentage of creak by time

Baseline voicing (s)	Creak (s)	Total duration	% creak by time
423.7	69.6	493.2	14.1

C.6.1 Data trimming

Following the same process as in the larger corpus, instances of constructed dialogue were excluded. This left 4298 stretches for analysis.

Duration and speech rate were scaled and log-transformed. No tokens occurred below the 1st quartile - 3*IQR or above the 3rd quartile + 3*IQR, so no tokens were discarded as outliers.

C.6.2 Statistical model

C.6.2.1 Model diagnostics

Model diagnostics followed the same procedures as in the larger corpus study. As shown in Figure C.1, the scaled residuals do not deviate significantly from uniformity, which is confirmed through Kolmogorov-Smirnov test (noted as KS test on the figure).

Figure C.1: Scaled (quantile) residuals for the mixed-effects logistic regression predicting creak

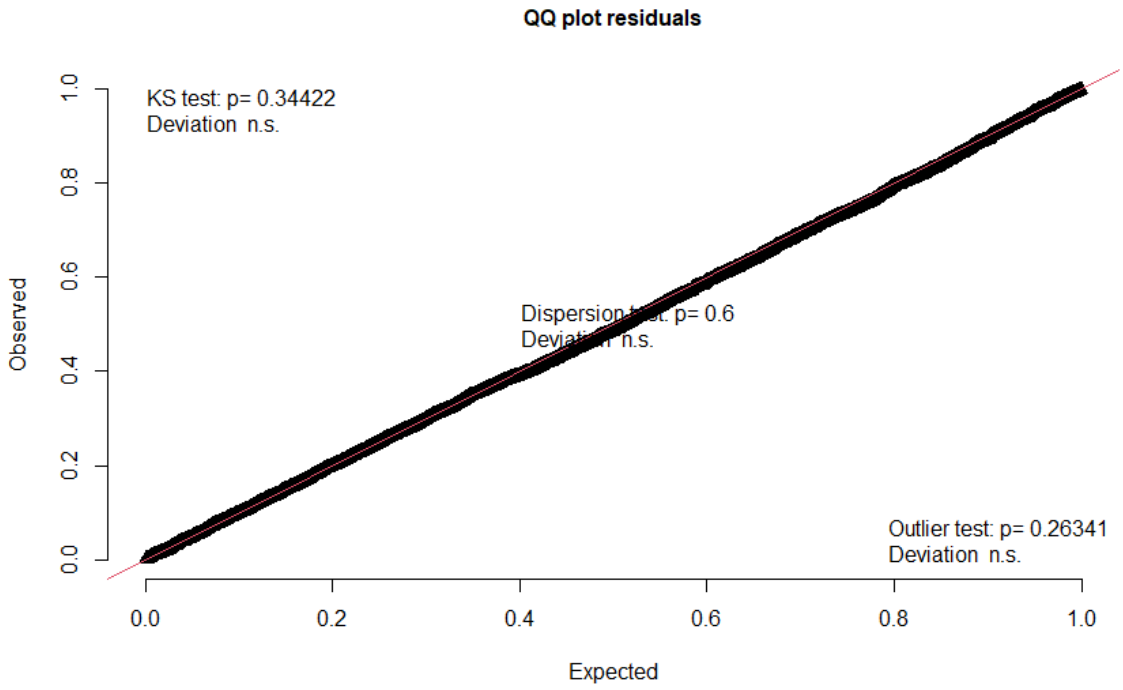


Table C.3: Pseudo R^2 measures for baseline and full model. Baseline model includes only a random intercept for words.

Model		R2m	R2c
Full	theoretical	0.34	0.47
	delta	0.23	0.32
Baseline	theoretical	0	0.18
	delta	0	0.10

There is no evidence of problematic collinearity in the model with all GVIFs < 1.5 and $c = 1.4$ for linear predictors.

Table C.3 shows pseudo-R^2 measures for the full and baseline model. This suggests more variance is explained in the full model.

Increased voiced stretch Duration and Speech Rate both reduced the log odds of creak (Duration $\beta = -1.130$, $SE(\beta) = 0.087$, $p < 0.001$; Speech Rate $\beta = -0.357$,$SE(\beta) = 0.061$, $p < 0.001$).

Creak was favoured by the potential presence of a glottal stop ($\beta = 0.862$, SE(β) = 0.117, $p < 0.001$), being followed by a potentil Preglottalisation context ($\beta = 0.696$, SE(β) = 0.284, $p < 0.05$), being in phrase Final position ($\beta = 0.434$, SE(β) = 0.116, $p < 0.001$), containing an Initial Vowel ($\beta = 1.031$, SE(β) = 0.175, $p < 0.001$), and containing no vowel ($\beta = 2.001$, SE(β) = 0.234, $p < 0.001$). On the other hand, containing a Nasal or a Rhotic decreases the log odds of creak (Nasal $\beta = -0.698$,

Table C.4: Results of mixed-effects logistic regression predicting the log-odds of creak in Carrie's speech as a function of linguistic factors

Independent variable	Level/unit	Coefficient (SE)
Intercept		-1.738^{***} (0.092)
		t = −18.964
Duration	Log transformed and scaled	-1.130^{***} (0.087)
		t = −12.954
Speech rate	Log transformed and scaled	-0.357^{***} (0.061)
		t = −5.889
Glottalisation	Glottalable context	0.862^{***} (0.117)
(Ref = No glottalisation)		t = 7.354
	Preglottalisation context	0.696^{*} (0.284)
		t = 2.449
Final	Final	0.434^{***} (0.116)
(*Ref = Not final*)		t = 3.733
Vowel	Initial vowel	1.031^{***} (0.175)
(*Ref = Non-initial vowel*)		t = 5.891
	Both	−0.687 (1.090)
		t = −0.630
	None	2.001^{***} (0.234)
		t = 8.555
Contains nasal	Nasal	-0.698^{***} (0.145)
(*Ref = No nasal*)		t = −4.817
Contains rhotic	Rhotic	-0.336^{*} (0.153)
(*Ref = No rhotic*)		t = −2.197
Observations	4,298	
Log Likelihood	−1,894.928	
Note:	*p<0.05; **p<0.01; ***p<0.001	

Table C.5: Random effects for the mixed-effects model presented in Table C.4

Groups	Name	Variance	Std.Dev.
words	(Intercept)	0.3356	0.5793
words.1	Duration	0.4621	0.6798

SE(β) = 0.145, $p < 0.001$; Rhotic $\beta = -0.336$, SE(β) = 0.153, $p < 0.05$/.

Table C.6: Results of the linear mixed-effect model predicting H2*–H4* as a function of linguistic factors in Carrie's voice

	Dependent variable:
	H2H4c_sc
Constant	0.240^{***} (0.042)
	t = 5.681
log_dur_scaled	0.134^{***} (0.023)
	t = 5.920
aspasp	−0.053 (0.044)
	t = −1.226
asph_asp	0.420^{***} (0.106)
	t = 3.967
contains_nasalnasal	-0.390^{***} (0.048)
	t = −8.129
contains_lwylwy	-0.137^{**} (0.046)
	t = −2.969
Observations	3,206
Log Likelihood	−4,410.408
Akaike Inf. Crit.	8,836.816
Bayesian Inf. Crit.	8,885.398
Note:	*p<0.05; **p<0.01; ***p<0.001

Table C.7: Results of the linear mixed-effect model predicting H4*–2kHz* as a function of linguistic factors in Carrie's voice

	Dependent variable:
	H42Kc_sc
Constant	0.028 (0.038)
	t = 0.722
log_dur_scaled	0.016 (0.024)
	t = 0.660
log_speech_rate_scaled	−0.095*** (0.021)
	t = −4.495
glottalisationglottalable_context	0.005 (0.053)
	t = 0.086
glottalisationpreglottalisation_context	−0.459*** (0.110)
	t = −4.182
aspasp	0.027 (0.043)
	t = 0.637
asph_asp	0.499*** (0.104)
	t = 4.808
vowelinitial_vowel	−0.058 (0.074)
	t = −0.779
vowelboth	−0.072 (0.151)
	t = −0.475
vowelnone	−0.409** (0.140)
	t = −2.919
contains_lwylwy	−0.157*** (0.046)
	t = −3.385
Observations	3,206
Log Likelihood	−4,052.514
Akaike Inf. Crit.	8,133.028
Bayesian Inf. Crit.	8,218.047
Note:	*p<0.05; **p<0.01; ***p<0.001

Appendix D

Online resources

D.1 Audio examples

Audio examples can be found in the online supplementary materials via OSF, but are not available openly due to the data originating from SCOSYA. To access them, please contact the author, Joe Pearce.

D.2 Praat scripts

List of scripts:

1. Extractlongeststretch3_directory.praat
2. Tier prep v4.praat
3. Extract by metadata.praat
4. cycle and exclude.praat

Praat scripts can be found in the online supplementary materials here: `https://osf.io/d2f6j/?view_only=6ff24b1962cb4a77a49e938b0b42fe56`