University of Sheffield

An Initial Investigation into the Effects of Facial Feminisation Surgery on the Voice

Cliodhna Hughes

Supervisors: Professor Guy Brown, Dr Ning Ma, Professor Nicola Dibben

A report submitted in partial fulfilment of the requirements
for the degree of Master of Philosophy in Computer Science

in the

School of Computer Science

March 2025

Abstract

Transfeminine people may choose to undergo facial feminisation surgery: an umbrella term covering a range of procedures that aim to alter the appearance of facial features, thereby potentially affecting characteristics of the vocal tract. However, effects of facial feminisation surgery on the voice are relatively understudied. This means that little information is available to people considering undergoing these procedures, about the potential impact on their voice. Such information is of particular relevance to people who use their voice professionally, such as singers. Previous work on the vocal effects of other facial surgeries also neglects to acknowledge the variety of factors that scholars in linguistics, and in particular sociolinguistics, have shown to affect the voice. This thesis presents an analysis of the effects of facial feminisation surgery on a transfeminine professional singer through a single case study. This task is approached from three different angles. First, the acoustic changes in the participant's voice are identified, through an acoustic analysis of speech and singing data collected from her before and after surgery. Then, the results of an experiment which aimed to identify the extent to which other people could perceive a change in her voice following the surgery are presented. Finally, the participant's own perception of the effects that the surgery had on her voice are explored, through a thematic analysis of longitudinal interview data. The results of the acoustic analysis and perceptual experiment suggest that facial feminisation surgery can have an impact on the voice, and the qualitative analysis suggests this may not only be as a result of the altered characteristics of the vocal tract, but also as a result of the altered social context. This research, then, provides an argument for including sociolinguistic methods in analysis of the impact of surgery on the voice.

Acknowledgements

I am grateful to so many people for helping me throughout this MPhil. First and foremost, I am thankful to my anonymous participant, who has given up so much of her time to help allow other people who are considering undergoing facial feminisation surgery to have more of an understanding of the potential vocal outcomes. I am grateful to her for trusting me with her voice data, and to tell her story. Getting to do this work has been a real privilege.

I am extremely grateful to my supervisors Guy Brown, Ning Ma and Nikki Dibben for their guidance, feedback, support, encouragement, and for having my best interests at heart. I also owe thanks to Emma Moore, Justin Christensen and Lauren Hall-Lew for helpful conversations around particular aspects of this thesis. Thank you to everyone in the Speech and Hearing Group and the Speech and Language Technology CDT, for providing a collaborative working environment, and from whom I have learned a lot. I am also hugely thankful to Stu Wrigley and Lizzie Pass, for answering my every question, sorting out all my admin problems and generally for helping me in so many ways, most of which I probably don't even know about.

I wormed my way into some other research groups and departments during this degree and I am grateful to have been welcomed by the people there. Muses Mind Machine were lovely to me throughout, and I enjoyed learning about music psychology research through their meetings. The Centre for Linguistics Research gave me a place to stay in touch with my linguistics roots, which I am very grateful for. I also owe thanks also to the Representation Theory group at the University of Cologne for giving me a desk and getting lunch with me while I was writing up.

I am thankful to all of my lovely friends, who are too many to name, for being so wonderful and never failing to cheer me up. Hannah, Harry, the Rams, Angus, Ed and Chan Dara in particular made this degree easier with their friendship, support, and silliness. I owe special thanks to the delightful Bea Livesey-Stephens for teaching me what a positionality statement is. I am also immensely grateful to Nina Markl and Annie Holtz, who are my friends but also my mentors, and who have always lifted me up and been so generous with their time, wisdom and kindness.

Finally, I am grateful to all my family for supporting me and listening to my fun facts about language throughout this degree, and last but not least to Max, for the endless encouragement and joy.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Trans people may feel that their face and/or voice do not align with their gender identity. For transfeminine people, facial feminisation surgery can help to ease this feeling with regards to the face, but it can also sometimes have an unwanted impact on the voice (Nuyen et al. 2023). There is relatively little work on this topic however, especially given the number and variety of surgeries included under the umbrella term of 'facial feminisation surgery.' Moreover, the work that does exist on this topic primarily comes from medical fields, where there is a tendency to view the voice as an artefact almost entirely determined by aspects of the speaker's biology (Azul 2013). Such a view disregards knowledge from sociolinguistics, which shows the voice to be highly flexible, and used by speakers to convey and construct different aspects of their identity in different ways in different contexts (Eckert 2012). This thesis provides an initial exploration of the effects of facial feminisation surgery on the voice, through a single case study, using approaches informed by both clinical fields and sociolinguistics. In this chapter, key terms are defined, the aims and limitations of the thesis are discussed, and my positionality with respect to this work is outlined.

1.1 Terminology

This thesis adopts a definition of *the voice* similar to the broad definition described by Kreiman et al. (2003): the voice is the combination of phonatory quality, articulatory details, pitch, amplitude variations and temporal patterning produced by a speaker, all of which contribute to conveying information about factors such as affect, attitude, pragmatics and, of most importance to this thesis, personal identity. The voice is used when producing speech, and when singing.

A person's *gender* is defined as the identity with which they feel internally aligned, which may be a term such as 'woman', 'man', 'non-binary', or another identity label (see e.g. Sanchez et al. 2024). This differs from *sex assigned at birth*, which is a categorisation made typically based on visible physical characteristics. *Perceived gender* is used to refer to the gender that a person is perceived as, or is likely to be perceived as, based on characteristics of their face or voice (as specified). In this thesis, the adjective *trans* is used to describe people whose gender does not align with their sex assigned at birth. This includes trans men and trans women, as well as people who do not identify with the labels man or woman. I use the term trans here because it is preferred by my participant, with the acknowledgement that individuals and communities will differ in their preferences for terms used to refer to them. In contrast, *cisgender* or *cis* describes people whose gender aligns with their sex assigned at birth. Gender and the related terms defined in this paragraph are primarily used in this thesis when talking about prior literature. However, as shown by Sanchez et al. (2024), such terms are often not defined in speech science papers. I have therefore occasionally made assumptions as to what is meant by the author(s) of discussed papers when using these terms, based on the content of the paper and the year of publication; it is important to note that these assumptions may not always be correct. Following Quinn & Hancock (2023), I acknowledge that preferred terminology will also change over time, such that the terms used here may become outdated.

The term *dysphoria* is used to denote the distress that a person may (but does not necessarily) feel when there is a mismatch between their gender identity and the gender assigned at birth (e.g. Coleman et al. 2022).[1] *Vocal dysphoria* is used to refer to this distress specifically when it stems from the voice. This can be because the person's voice does not align with expectations of how a voice of someone of their gender 'should' sound. Such expectations are socially driven, and may be those of the person who is experiencing the dysphoria, or those of people around them.

1.2 Aims

For trans people, this thesis aims to add to what is known about how the voice can be affected by facial feminisation surgery. The primary research question that runs throughout the experimental and data analysis chapters is: what impact can facial feminisation surgery have on the voice? This is addressed in chapters 4, 5 & 6, via three smaller-scale research questions:

RQ1 How can facial feminisation surgery impact the voice acoustically?

RQ2 Are acoustic changes to the voice following facial feminisation surgery perceptually noticeable to external listeners?

RQ3 How can facial feminisation surgery affect a person's own perception of, and relationship with, their voice?

For speech scientists who work with trans people, and with other people who may be undergoing facial surgeries, this thesis aims to highlight the inadequacy of the methods used in previous work to evaluate the impact of facial surgeries on the voice, and to draw attention to the lack of information available on the potential impact of facial feminisation surgery on the voice. These aims are addressed in chapters 2 & 7, as well as in the background sections of each of chapters 4, 5 & 6.

1.3 Limitations

To prevent this work being misinterpreted, here I list some key limitations and make explicit what this research is not.

[1] Dysphoria is primarily discussed in Chapter 6, which provides an analysis of interviews in which the participant discusses her experience of the phenomenon. The choice to adopt this definition stems from the participant's own use of the word.

Firstly, the results in this thesis are based on a single case study. Making speech, singing and interview recordings at several points in a person's transition journey requires a substantial amount of time commitment at a potentially very emotionally demanding period in the participant's life. Additionally, many trans people experience vocal dysphoria before receiving voice-altering healthcare (e.g. hormones, gender affirming speech therapy), which can make them less likely to want to record their voice and consent to these recordings being used for research purposes, at least at particular stages in their transition journey. Because of these difficulties in obtaining this kind of data, there is value in doing this type of research with only one participant. It also allows a greater depth of insight than would be possible with a larger scale study. However, having only one participant means that this work simply provides an example of how one person's voice was impacted by facial feminisation surgery. Different people will experience different vocal impacts. This thesis does not present a representative or exhaustive description of the ways in which facial feminisation surgery could affect the voice.

This issue is compounded by the fact that facial feminisation surgery is an umbrella term covering a number of surgeries, which makes it is extremely difficult to make links between specific vocal tract alterations and speech changes. This thesis does not attempt to provide a list of vocal changes which correspond to specific surgical alterations of the face, rather it gives a broad overview of the changes that occurred for this particular participant with the surgeries that she underwent.

Finally, while I argue that prior methods for analysing the effects of facial feminisation surgery are inadequate, and suggest some alternatives, I do not argue that future studies should be based on the methods used in Studies 1, 2 & 3 of this thesis. The work conducted for this thesis was done using data collected for a different purpose, as will be explained in chapter 3, and therefore possesses a number of further limitations, which are discussed in greater depth in chapters 4, 5 & 6.

1.4 Positionality

Creation and analysis of data is reflexive, in that the experiences and beliefs of researchers affect how they approach research. This fact is most often acknowledged in qualitative research, but is true of all research. For this reason, I outline aspects of my identity and background below, for transparency and to contextualise my choice of methods and the conclusions I draw.[2] I am grateful to Quinn & Hancock (2023) for their guidance on writing a positionality statement in papers on speech science research with trans people.

I approach this work as a queer cis woman and ally to trans people. I hold the view that gender is a social construct, is fluid and intersectional, and is not determined by biological characteristics. I currently see the voice both as a highly flexible artifact, that people can (learn to) use in vastly different ways (as indicated by sociolinguistics and speech therapy), as well as something which is at least influenced by aspects of a speaker's biology (as indicated by the experiences of trans people). I recognise that I come from a place of privilege and believe that research with marginalised communities should centre around their needs and desires, and so I aim to allow my research aims and methods to be guided by the participants

[2]Note that the majority of the data was collected by other members of the research team, and this positionality statement applies only to me, as the author of this thesis and the person who analysed the data.

I work with. I am learning to recognise when my own career goals or research interests might be impeding this. In the case of the present work, the participant approached my supervisors and identified that there was a lack of information available to her regarding the potential vocal impact of facial feminisation surgery, so this research was motivated by a community need. My own desire to be involved in this project stemmed from my ongoing beliefs that gender-affirming healthcare is important, and that so too is the ability for a person to make informed decisions around their healthcare. My formal training before engaging in the present work came primarily from the fields of linguistics and speech technology, where I admittedly learned little about conducting qualitative or social research. My training in and love for phonetics, in particular, motivated my choice of approach for Study 1 of this thesis, where I initially aimed to simply provide an 'objective' account of the acoustic changes to the participant's voice before and after undergoing facial feminisation surgery. When I could not explain the observed results straightforwardly in terms of the changing shape of the participant's vocal tract, I was forced to consider the broader social context, leading to Study 3. The work in this thesis forms my first experience of conducting research with and for trans people, and Study 3 was my first experience of conducting qualitative research, so I still have much to learn.

1.5 Thesis Overview

The question of how the voice can be affected by facial feminisation surgery is approached from three different angles in this thesis. In chapter 2, relevant background literature is presented. Chapter 3 introduces the dataset used in the rest of the thesis. In chapter 4 an acoustic analysis of recordings of speech and singing data taken from the participant before and after undergoing facial feminisation surgery is reported. Chapter 5 details a perceptual experiment aimed at investigating whether the differences in the participant's pre-surgery and post-surgery recordings are perceptually salient. Chapter 6 presents a qualitative analysis of interviews conducted with the participant before and after undergoing facial feminisation surgery. In chapter 7, I discuss the findings of the three preceding chapters and spell out the implications of this research, before summarising the conclusions that can be drawn.

Chapter 2

Background

2.1 The Voice and Identity

The voice is fundamentally linked to the speaker that produces it, in that it consists of features that reflect aspects of the speaker's physiology, and index aspects of the speaker's identity. On a basic level, vocal production involves air from the lungs causing the vocal folds to vibrate, creating sound which is shaped by the vocal tract before leaving the mouth as speech or song. In this way, the voice is affected by the shape and size of the vocal tract, and properties of the vocal folds. The shape and size of the vocal tract and nature of the vocal folds are in part influenced by anatomy (e.g. taller people tend to have longer vocal tracts; Fitch & Giedd 1999), but can also be influenced by the speaker to a large extent (e.g. through manipulation of the articulators). Sociolinguists have shown that speakers adapt features of their voice both consciously and subconsciously to index specific aspects of their identity in different ways in different social contexts. Over time, linguists' understanding of the relative contributions of biological and social factors on the voice has changed substantially.

Historically, linguists have often assumed observed differences between genders in speech and singing to be due to biological differences, despite a lack of direct evidence for this (see Zimman 2018, for a full discussion). One example of a feature of the voice which is often stated to be physiologically determined is the average fundamental frequency (F0) (see e.g. Ohala 1984). Cadaver studies show that adults assigned male at birth generally have larger and thicker vocal folds than adults assigned female at birth, and computational simulations suggest that this is a contributing factor to the lower average F0 observed in cisgender men compared with cisgender women, since weightier vocal folds vibrate more slowly (Titze 1989). Research on transmasculine people undergoing hormone therapy also provides evidence in support of this, as average F0 tends to decrease as time on testosterone increases (e.g. Zimman 2017). Physiological factors therefore certainly contribute to a person's average F0. A frequently stated explanation for the drop in average F0 in transmasculine people undergoing hormone therapy is that the exogenous testosterone causes physical changes in the vocal tract similar to those which occur during puberty for cisgender men, though as Azul (2015) points out, there is a lack of empirical studies to verify this. Additionally, as noted by Azul (2013) and Zimman (2017, 2018), among others, there is plentiful evidence that speakers' F0 patterns are not purely biologically determined. For example, there are reported differences between different languages and cultures, as well as across time, in average F0. Traunmüller

& Eriksson (1995) brings together a number of published mean F0 values across different languages, varieties and speech styles, finding substantial differences between them. Yuasa (2008) finds that speakers of Japanese display a greater distance between men and women in their average F0 than American English speakers, and that average F0 results for American English speakers are lower than those of demographically similar speakers from the 1950s and 1960s. Additionally, Szakay (2006) finds significant differences between Maori and Pakeha (white) speakers of New Zealand English in their average F0. These results demonstrate that social factors have a great deal of influence on average F0, and that it is therefore not purely determined by a speaker's physiology.

Going beyond F0, various sociolinguistic studies have shown that people use a variety of linguistic features both consciously and subconsciously, to index specific aspects of their identity, including gender, sexuality, socioeconomic background, race, and geographical ties in speech (see e.g. Labov 1966, Omoniyi & White 2006, Zimman 2017, Eckert & McConnell-Ginet 2003) and in singing (see e.g. Gibson 2019, Papineau 2020). As a result, in modern sociolinguistics, variation in speech is generally regarded as something which both reflects and constructs social meaning (Eckert 2012). People are viewed as having a great deal of autonomy and control in the way that they speak and sing, and are seen to speak and sing in different ways in different contexts to communicate different aspects of their identity to the people around them.

The view that the voice is largely determined by biology is neither exclusive to older research nor to linguistics research: several other fields show evidence of having this view. For example, the development of speaker recognition and verification models and their deployment in contexts such as biometric authentication in banking indicates a belief that the voice is unique to individuals, similar to a fingerprint in the way that it reflects a person's vocal tract. In medical research, too, there is a historical tendency to attribute a substantial amount of the variation between voices to differences in people's physiology. As discussed by Azul (2013), evidence used in the medical literature on the differences between the voices of people who differ in their sex assigned at birth tends to come from cadaver studies, computer tomography and acoustic reflection studies. Such methods provide evidence for anatomical differences between different groups, but reliance purely on these methods overlooks the fact that different people make use of their anatomical vocal apparatus in different ways. Speech therapy provides one example of a medical subfield in which this view is changing: recent years have seen an increase in the research and resources available on voice training for trans people (e.g. Hancock & Helenius 2012, Adler et al. 2019, Buckley et al. 2020, Mills et al. 2021). At a minimum, the existence of such literature conveys the view that a person has the potential to alter their voice (at least to an extent) with conscious effort and training, given their physical vocal tract.

However, the surgical literature continues to paint a different picture. Recent studies on the effects of various surgeries on the voice persist in reflecting the belief that the voice is primarily determined by anatomy. Studies on the vocal impact of rhinoplasty, for example, often take acoustic measurements from nasal consonants and vowels uttered in isolation before and after surgery (e.g. Guarro et al. 2019, Bakhshaee et al. 2021). The use of such methods is indicative of a belief that the only factor that could affect a person's voice in these conditions is the altered shape of the vocal tract, as caused by the surgery, and that any potential that an individual has to consciously or subconsciously affect the measurements taken is either non-existent or negligible.

2.2 Speech, Gender and Trans Voices

Trans people can provide a great deal of insight into the impact of physiology on the voice, as well as the extent of control that a speaker has over particular features of their own voice. As mentioned above, specific vocal traits are culturally associated with specific genders. As well as the aforementioned F0, centre of gravity of /s/, and formant spacing (a correlate of the overall resonant properties of the voice) also contribute to a speaker's perceived gender (e.g. Cartei et al. 2014). As discussed, physiology contributes in part to F0, but it can also be modified by the speaker. Perceptually, listeners associate a higher centre of gravity in /s/ productions with the speech of women (e.g. Strand 1999), as well as that of gay men (e.g. Linville 1998). Centre of gravity of /s/ is correlated with the position of the tongue in relation to the alveolar ridge, with more fronted productions causing a higher centre of gravity (associated with feminine speakers and gay men), and more retracted tongue positions causing a lower centre of gravity (associated with heterosexual masculine speakers). Centre of gravity of /s/ is therefore a feature that can be consciously changed by the speaker, and the observed differences between genders are generally agreed to be entirely socially motivated. There is some evidence that the size and shape of the vocal tract contributes to formant spacing, with longer vocal tracts resulting in a lower average distance between formants (e.g. Cartei et al. 2014), but formants are also related to the placement of the articulators, and can therefore be manipulated by the speaker.

Since specific vocal traits are associated with specific genders, the voice is often a source of dysphoria for trans people (e.g. Hancock et al. 2011), and surveys indicate that many seek out healthcare options for changing their voice (e.g. Eyssel et al. 2017). There are currently three main categories of gender-affirming healthcare options for changing the voice: hormone therapy, voice and communication therapy, and laryngeal surgery (Coleman et al. 2022).

Changes in hormones can impact the voice, and there is some evidence for vocal changes throughout the menstrual cycle and with menopause (see e.g. Silverman & Zimmer 1978, Boulet & Oddens 1996, Çelik et al. 2013, Lã & Ardura 2022). For trans people, exogenous oestrogen generally has little to no effect on the voice (Mészáros et al. 2005). Exogenous testosterone usually results in a decrease in average F0, and there is also some evidence that it causes some changes in the formants, indicative of a larger vocal tract length (e.g. Deuster et al. 2016, Hancock et al. 2017, Hodges-Simeon et al. 2021). It can also have undesired effects on the voice, however, such as a more restricted pitch range, and vocal instability (e.g. Coleman et al. 2022, Graham 2022).

Voice and communication therapy can help trans people acquire a voice that better fits with their gender identity. Such voice training typically focuses on adapting the F0 range, formants, intonation patterns and voice quality (e.g. Hancock & Helenius 2012, Adler et al. 2019, Mills et al. 2021). Transmasculine people have generally received little focus in the voice and communication therapy literature (Azul 2016), and their uptake in voice and communication therapy is lower than that of transfeminine people (Ziltzer et al. 2023). This is possibly due to the imbalance in the extent of the effects that exogenous testosterone and oestrogen have on the voice, since there is evidence for a greater improvement in self-perception of the voice after hormone therapy (with no corresponding voice training) for trans men than for trans women (Bultynck et al. 2017). However, as recent research has shown, hormone therapy alone is often not sufficient in helping transmasculine people to reach their vocal goals. Azul (2015) outlines some reasons for this. Firstly, there are several linguistic features other than F0 which are associated with particular genders, and which are not expected to change with hormone therapy (e.g. intonational patterns, centre of gravity of /s/). Secondly, testosterone does not have the same effect on the voices of all transmasculine people: the extent of F0 decrease differs considerably between transmasculine people undergoing hormone therapy. Finally, the extent to which the person undergoing hormone therapy is satisfied with their new voice, and the extent to which their voice is perceived as masculine by themselves or others can also differ greatly between people (e.g. Van Borsel et al. 2000, Zimman 2017). In recent years, there has therefore been a gradual increase in research and resources surrounding voice and communication therapy for transmasculine people (e.g. Adler et al. 2019, Buckley et al. 2020).

There are a number of pitch-lowering (e.g. thyroplasty type III, vocal fold injection augmentation) and pitch-raising (e.g. glottoplasty with retro-displacement of the anterior commissure, cricothyroid approximation, feminization laryngoplasty, laser-assisted voice adjustment) surgeries available to trans people (Coleman et al. 2022). There is limited research on pitch-lowering surgeries, but recent studies indicate they are successful in lowering mean F0 (Bultynck et al. 2021, Webb et al. 2022). Pitch-raising surgeries are reported to increase mean and minimum F0 (Kelly et al. 2019, Coleman et al. 2022). However, a number of possible negative side effects of pitch-raising surgery have been reported, such as reduced amplitude, decreased vocal range, pitch instability and hoarseness (Song & Jiang 2017, Kelly et al. 2019, Coleman et al. 2022).

2.3 Facial Feminisation Surgery

The face poses another possible source of dysphoria for trans people, since specific facial characteristics are also culturally associated with particular genders. For transfeminine people, facial feminisation surgery can help to ease this dysphoria. Facial feminisation surgery is an umbrella term for a range of procedures for changing the appearance of the facial features (Altman 2012). These surgeries do not aim to change the voice, but since they involve altering properties of the vocal tract, vocal changes can occur as a side-effect (e.g. Nuyen et al. 2023). In this section, four types of facial feminisation surgery are discussed: cranioplasty, which targets the forehead; rhinoplasty, which targets the nose; genioplasty, which targets the chin; and chondrolaryngoplasty, which targets the neck. Figure 2.1 shows this information graphically. For each surgery type, a brief lay description of the procedure is given, and the potential and/or observed effects on the voice are described.

2.3.1 Cranioplasty

The shape of the forehead has been shown to contribute to femininity ratings (Spiegel 2011). As a result, a number of surgical procedures exist for feminising the forehead, with different techniques used depending on the face shape of the person undergoing the surgery (Altman 2012). In general, these surgeries involve removal of bone, and occasionally addition of bone cement, in the forehead (Altman 2018). Since the forehead is not a component of the vocal

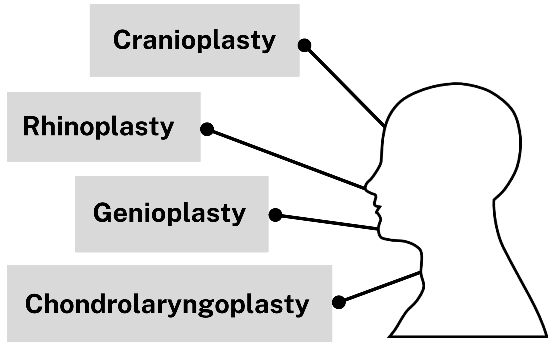

Figure 2.1: *Diagram showing the areas targeted by four types of facial feminisation surgery.*

tract, this type of facial feminisation surgery is not expected to affect the acoustics of the voice, and indeed there seems to be no research as of yet on the impact of these forehead surgeries on the voice. However, singers often report ‘feeling’ the vibrations of their singing voice in locations on the body such as the forehead, and so this may be affected by the altered characteristics of the forehead brought on by facial feminisation surgery. A change in the sensory feedback this provides may indirectly affect a person’s singing ability, making it more difficult for them to gauge whether they are achieving particular vocal qualities (see e.g. Brown 2002).

2.3.2 Rhinoplasty

Rhinoplasty involves altering the shape of the nose. As well as being an instance of facial feminisation surgery, it is also a procedure which is commonly undergone by cisgender people, both for cosmetic reasons and to improve breathing (see e.g. Rohrich & Ahmad 2011). It can involve addition or removal of bone or cartilage in the nose, depending on the requirements of the person undergoing the surgery (e.g. Daniel 2018). As the nasal cavity forms part of the vocal tract, rhinoplasty can have an impact on the acoustic properties of the voice. Several researchers have investigated the vocal impact of rhinoplasty, on presumably cisgender participants (e.g. Foroughian et al. 2014, Guarro et al. 2019, Xiao et al. 2020, Bakhshaee et al. 2021, Claros et al. 2021). Results suggest that rhinoplasty impacts nasal phones, causing an increase in frequency and a decrease in amplitude of nasal murmurs (Foroughian et al. 2014, Nemati et al. 2019). In general, however, these studies display a number of methodological issues, as described in section 2.1, and should therefore be considered with caution.

2.3.3 Genioplasty

Genioplasty involves repositioning or reshaping the chin, which can be done through addition or removal of bone (see e.g. Gursky et al. 2024). Like for cranioplasty, there seems to be no published research on the potential impact of genioplasty on the voice as of yet. However, since the chin forms part of the jaw, it is a component of the vocal tract and it is therefore possible that genioplasty could have an impact on the voice, for example through altering the characteristics of a component of the vocal tract wall, which is known to affect formant frequencies and bandwidths (Fleischer et al. 2015).

2.3.4 Chondrolaryngoplasty

One further instance of facial feminisation surgery is chondrolaryngoplasty, also known as thyroid shave, which involves removing some of the thyroid cartilage, thereby altering the shape of the neck (Altman 2012). This procedure has been reported to sometimes lead to vocal changes which are generally undesired by transfeminine people, such as hoarseness, lower average F0 and lower maximum F0 (e.g. Strickland et al. 2022, Nuyen et al. 2023). These negative vocal effects are not always observed (see e.g. Aires et al. 2021), and seem to generally be caused by surgical errors, which can sometimes be fixed with further surgery (Nuyen et al. 2023). It is possible that successful chondrolaryngoplasty could also have an impact on the voice, since it involves removing material close to the vocal tract, but there seems to have been no investigation of this possibility yet.

2.4 Summary

In general, there is a scarcity of research on the impact of facial feminisation surgery on the voice, at least for the four types of facial feminisation surgery discussed here. The studies that have been done focus exclusively on the effects of rhinoplasty and chondrolaryngoplasty on the voice. As well as displaying a number of methodological issues, the studies on rhinoplasty have seemingly only been done with cisgender participants. Chondrolaryngoplasty has gained attention because of its potential to have severe negative side effects on the voice, and so the potential effects of a successful surgery on the voice have yet to be investigated. Other facial feminisation surgeries such as cranioplasty and genioplasty also have the potential to affect the voice, through altering characteristics of the vocal tract, as well as areas in which singers can feel the vibrations of their voice when singing.

Additionally, the studies that have investigated the vocal impact of rhinoplasty and chondrolaryngoplasty have done so using methods that disregard the influence of social factors on the voice, and the speaker's own capacity to consciously or subconsciously alter their voice. Such medical research is influential in the lives of trans people, since it provides them with information upon which to base their decisions about their own healthcare. However, the surgical literature in particular has yet to catch up with the modern understanding of the complex relationship between physiology, gender and the voice, developed in sociolinguistics, and by trans people themselves (e.g. Azul 2013, Zimman 2018). In the case of facial feminisation surgery, this is of particular concern for two reasons. Firstly, the visible facial changes that the surgery causes can impact how the speaker is perceived by other people, thereby altering the social context for any interaction, which is known to affect speech. Secondly, as part of their transition, it is likely that the speaker could be undergoing gender-affirming speech therapy or other formal or informal types of voice training between recording sessions.

The lack of research in this area means that there is a scarcity of information available to trans people considering undergoing facial feminisation surgery on the potential impact it may have on their voice. This information is of particular consequence to people who use their voice professionally, such as singers, actors, teachers and radio hosts, to name but a few. This thesis therefore aims to provide an initial investigation into the effects of facial feminisation surgery on the voice, through analysis of data from one participant. The following chapter describes this data.

Chapter 3

Data

The data used in chapters 4-6 of this thesis consists of speech, singing and interview recordings from a transfeminine singer who underwent the four types of facial feminisation surgery outlined in section 2.3. Since the same data is used in all three Studies in this thesis, a detailed overview of the data is provided in this chapter. Importantly, the data were not collected for the purpose of analysing the impact of facial feminisation surgery on the voice. The participant was concerned that the surgery would have a permanent and negative effect on her voice, and so the speech and singing recordings were collected for the purpose of building a voice conversion system that would allow her to convert from her post-surgery voice back to her pre-surgery voice. Her initial intention was to use this creatively in performance. However, due to several factors, including the singer's changing relationship with her voice, the decision was made not to build a singing voice conversion system and to instead use the data to document the acoustic changes to the voice caused by surgery. This reflects an important aspect of participatory research, which is that it requires a willingness to adapt the methods and aims of the research according to the needs of the participants.

I joined the project after most of the data had already been collected, and so I had no involvement in the collection of speech, singing or interview data, with the exception of the eleven months post-surgery interview and the validation interview, which I conducted.

3.1 Participant

The participant, who is an adult trans woman, is a professional singer based in the United Kingdom. Here, she is referred to by the pseudonym Maria. Maria underwent facial feminisation surgery after the first recording session. Specifically, the procedures were cranioplasty, rhinoplasty, genioplasty and chondrolaryngoplasty. Her first language is a Slavic language, and she describes her English accent as 'international', having been influenced by American and British media.[1] Maria was 25 years old at the time of surgery. She has been singing since around age six or seven, and has extensive vocal training that she describes as stretching from "traditional and choral singing" to "more contemporary, pop, solo performance." She performs both as a solo artist, and in two choirs. At the beginning of the data collection period, she positioned herself as a bass-baritone when singing in a group professionally. She describes her solo practice as being quite commercially-oriented, and lists Troye Sivan and

[1] All research for this thesis was conducted in English.

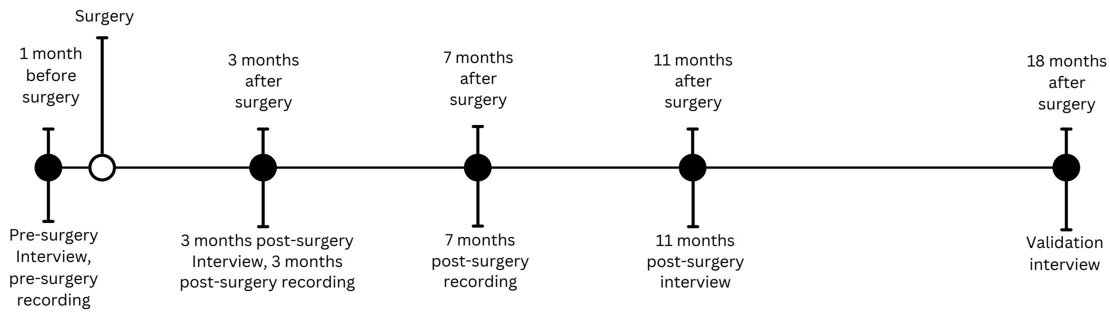

Figure 3.1: *Data collection timeline, with time points given in relation to the surgery date.*

Caroline Polachek as inspirations. In the year prior to surgery, she had four gender-affirming speech therapy sessions.

3.2 Timeline

Speech and singing recordings, which are used in Studies 1 and 2, were collected at three time points: one month prior to surgery, three months post-surgery and seven months post-surgery. Interviews for Study 3 were conducted at three different time points: one month prior to surgery, three months post-surgery and eleven months post-surgery. Additionally, a validation interview was conducted 18 months after surgery. These timepoints were chosen somewhat arbitrarily, and were not based on prior literature. Figure 3.1 is a timeline, showing the dates of interviews and speech and singing recordings.

3.3 Speech and Singing Recordings

Singing and speech recordings were made in Maria's recording studio using a Sonotronics Aria microphone with a UA-610B (UAD plugin) pre-amp, with a 48kHz sampling rate and a 24-bit depth. For the speech data, Maria was asked to read the Rainbow Passage at each of the three time points (Fairbanks 1960). The singing material consisted of a subset of songs from the NUS Sung and Spoken Lyrics Corpus (Duan et al. 2013). Only songs which were recorded in all three sessions were used in this thesis, so that the phonemic material was consistent across sessions. Specifically, Maria's renditions of *Can You Feel the Love Tonight*, *I Have A Dream*, *Jingle Bells*, *Moon River*, *Silent Night*, and *You Are My Sunshine* were used in Studies 1 and 2. In total, there was 2 minutes of speech and 16 minutes of (analysable) singing data on average from each of the three recording sessions.

3.4 Interview Recordings

Various members of the research team conducted interviews with Maria, to track changes in her feelings and perception regarding her voice. Employing a semi-structured interview approach, the research team prepared a list of questions for Maria before each interview. These explored Maria's rationale behind opting for facial feminisation surgery and engaging in this research, alongside her perception of her voice. Maria was encouraged to also share other aspects of her experience that she felt were relevant to the project, and in this way she described her experiences concerning the interplay between her voice, physical appearance, and sense of self-identity, as well as her anticipations and concerns regarding both her voice and identity. Since the original purpose of this research was for building a singing voice conversion system, and because Maria was primarily concerned with the potential impact of the surgery on her singing voice, the first two interviews focused primarily on her singing voice, and her speech was seldom mentioned. The data analysed in Study 3 (Chapter 6) comes from three interviews. The first was conducted one month before surgery, the second three months after surgery, and the third eleven months after surgery.

Since Study 3 involved drawing conclusions about a particular person's thoughts and feelings, participant validation was used to ensure the conclusions drawn accurately reflected Maria's experience. After I had defined and named the themes, I conducted a validation interview with Maria. Three other members of the research team were also present. In this interview, I presented each of the themes to her, asked for her input on whether these accurately represented her thoughts and feelings at the time of the original interviews. We discussed each theme and the topics within it individually, as well as whether they altogether painted an accurate overall picture. Following this, I amended some of the specifics of particular themes, in line with Maria's comments.[2] The validation interview took place 18 months after the surgery. All four interviews were conducted and recorded on Google Meet.

3.5 Ethics

This research was granted ethical approval by the Ethics Committee of the the Music Department at the University of Sheffield. Maria approached the research team with a project in mind, and so took part in the research actively, and with interest. The aims of the research changed in conversation with Maria. As new research questions were defined and updated methods were proposed, amendments to the original ethics application were submitted and approved, and Maria gave informed consent for her data to be used in these new ways. Data were stored securely, in accordance with best practices.

[2] Insights coming specifically from the validation interview are marked as such in the text in Chapter 6.

Chapter 4

Study 1: Acoustic Effects

4.1 Introduction

Acoustic analyses are frequently used to investigate vocal changes caused by voice-altering gender-affirming healthcare interventions such as hormone therapy, pitch raising or lowering surgery and speech therapy. Such methods give an insight into whether any physiological and/or articulatory changes have taken place, and provide an indication as to the extent of them if so. This chapter describes an acoustic analysis of speech and singing data taken from Maria before and after undergoing facial feminisation surgery. It is an extended version of a paper presented at Interspeech 2024 (Hughes et al. 2024).

4.2 Background

As outlined in section 2.3, facial feminisation surgery has the potential to affect a number of acoustic variables. Specifically, through altering characteristics of the vocal tract, the surgery could affect the resonances, and chondrolaryngoplasty specifically is known to also sometimes impact voice quality. For this reason, this study analyses changes in resonances as measured by formants F1, F2 and F3, and in voice quality as measured through long-term average spectra (LTAS) and H1-H2, H1-A1, H1-A2 and H1-A3. Additionally, centre of gravity of /s/ measures are taken as a control. Since this is a variable that is correlated with gender, and this variation is believed to be entirely socially motivated (i.e. does not have any anatomical explanation), it is not expected to change as a direct result of the surgery altering the shape of the vocal tract. Any observed differences in centre of gravity of /s/ before and after surgery can therefore be attributed to (conscious or subconscious) articulatory changes made by the speaker. Formants and voice quality measures are acoustic variables which are expected to be affected by physical properties of the vocal tract, but can also be modified with (conscious or subconscious) control from the speaker. If changes in centre of gravity of /s/ are observed, then this indicates that Maria is making articulatory changes to her voice, and suggests that any observed formant and voice quality changes may not be entirely attributable to the physical changes to the vocal tract that the surgery causes[1].

[1] This is of course not a watertight method: a person can consciously or subconsciously change their formants or voice quality independently of centre of gravity of /s/.

4.2.1 Formants

Formants are frequency regions in which there is relatively more acoustic energy than in other areas. In speech and singing, they are caused by the voice source passing through the vocal tract, which acts as a resonator (see e.g. Sundberg & Sataloff 2005, for an introduction). They are therefore affected by the size and shape of the vocal tract, which can change as a result of surgery, and as a result of the speaker moving their articulators (see e.g. Story et al. 2001). In speech, formants cue the listener as to the phoneme being produced by the speaker, and are particularly relevant for vowel perception, with the first formant (F1) being correlated with tongue height and and the second formant (F2) being correlated with tongue backness in vowel production (Delattre 1951, Stevens & House 1955). Formants also provide the listener with some limited physiological information about the speaker, as formant patterns tend to be correlated with speaker height (Pisanski et al. 2014), though this is not straightforwardly interpreted by listeners (see e.g. Barreda 2016). Relatedly, formants also contribute to gender perception (e.g. Schwartz & Rine 1968, Cartei et al. 2014). In the cases of height and gender perception, the precise relationship between formants and height or gender is not clear, with some studies finding the relationship to be based on average formant frequency, some finding it to be based on distance between formants and others on other patterns between formants (e.g. Pisanski et al. 2014).

4.2.2 Voice Quality Measures

The relationship between perceived voice qualities, articulation and acoustics is not as well understood as that between formants and articulation. Additionally, no one acoustic measure distinguishes between all voice qualities (see e.g. Esposito 2012, Esposito & Khan 2020). As a result, many acoustic measures exist for quantifying differences in voice qualities. This study therefore takes a variety of voice quality measurements from the data. Specifically, measurements which normalise the amplitude of the first harmonic (H1) against other spectral landmarks, and Long Term Average Spectra. Like formants, voice quality can be affected by both physiological factors and articulatory decisions. The potential influence of physiology on voice quality is evidenced by the aforementioned reported hoarseness in some people who have undergone chondrolaryngoplasty (e.g. Nuyen et al. 2023). The fact that voice quality is phonemic in some languages (e.g. breathy vowels in Gujarati Fischer-Jørgensen 1967) provides one example of evidence for the fact that speakers have some articulatory control over their voice quality. The most commonly reported gender difference in voice quality is breathiness being associated with femininity (e.g. Henton & Bladon 1985, Mendoza et al. 1996, Whitling et al. 2023). This is typically believed to be a socially motivated difference, though some have argued that it is in part caused by the thinner and lighter vocal folds that cisgender women tend to have.

H1-related Measures

In an analysis of breathy vowels in Gujarati, Fischer-Jørgensen (1967) found that vowels with the breathy voice quality exhibited an amplified first harmonic (H1), compared with vowels produced with modal voice. Since overall loudness of the voice can also affect amplitude of H1, the decision was made to normalise it against another spectral landmark, as an increase in amplitude should affect each spectral landmark in the same way. Specifically, the second harmonic (H2), and the amplitude of the first, second, third and fourth formants (A1, A2, A3, and A4, respectively), were candidates. Fischer-Jørgensen (1967) found that H1-H2, H1-A1, H1-A2 and H1-A4 all distinguished between breathy and modal vowels in her data, but that H1-A3 did not. Subsequent research aimed to investigate the relationship between H1-H2 and production, with inconsistent results. Kreiman et al. (2012) found that different speakers used different articulatory strategies to manipulate H1-H2, indicating that it does not have a one-to-one correspondence with a particular articulatory movement, leading them to conclude instead that different speakers use their vocal apparatus in different ways to achieve particula voice qualities. Other researchers have found H1-related measures to correspond with perceptual voice quality descriptors other than breathiness. For example, Narasimhan & Vishal (2017) find H1-H2 and H1-A1 to correlate with hoarseness. Gender differences in H1-related measurements have also been reported, with Henton & Bladon (1985) finding the British English (presumably cisgender) women in their sample to have higher H1-H2 than the (presumably cisgender) men in their sample. This corresponds to the women in their sample speaking with a breathier voice quality, which the authors believe to be socially motivated.

Long Term Average Spectrum

A Long Term Average Spectrum (LTAS) provides an average of the spectral characteristics of speech over an extended period of time, such as a read passage. It is intended to average over the phonetic details of individual segments and therefore provide information on the sound source (Löfqvist 1986); Li et al. (1969) suggest that after 30 seconds of continuous speech the shape of the LTAS becomes stable. There has been much research into its utility in diagnosis of vocal pathologies and assessment of vocal healthcare outcomes, with Tanner et al. (2005) finding that moments derived from the LTAS correlated with perceived improvement in dysphonia severity following speech therapy, for example. Linville (2002) finds that the LTAS also changes with age, which is believed to be due to physiological changes in the vocal tract which occur as part of the ageing process (Hirano et al. 1989). Mendoza et al. (1996) investigate differences in the LTAS of people who differ in their sex assigned at birth. They find that the LTAS of people assigned female at birth tends to contain more energy in the higher frequencies, which again corresponds to a perceptual quality of breathiness.

4.2.3 Centre of Gravity of /s/

Production of the fricative [s] involves forming a constriction between the tongue blade or tip, and the alveolar ridge (e.g. Dart 1991). A more fronted production (i.e. where the tongue tip is closer to the teeth) results in the mean of the acoustic energy across the frequency spectrum (the centre of gravity) being higher (e.g. Jongman et al. 2000). A fronted production and higher centre of gravity has been repeatedly found to be a feature of the speech of women in production in English (Jongman et al. 2000, Fox & Nissen 2005). In perception, it is a feature which has been shown to affect perceived gender as a category, as well as masculinity in English speakers (Avery & Liss 1996, Strand 1999). Centre of gravity of /s/ measurements are also correlated with sexuality, with /s/ productions of gay masculine speakers having a higher centre of gravity in English (Linville 1998). This has been shown to be salient in perception as well, with centre of gravity playing a role in listeners' perception of speaker sexuality (e.g. Boyd 2018). Due to the relationship between centre of gravity of /s/ and production, as well as the fact that it is correlated with sexuality, the variation in centre of gravity of /s/ between genders is believed to be purely socially, rather than biologically motivated.

4.2.4 Summary

Facial feminisation surgery has the potential to affect a number of acoustic variables through altering the physical properties of the vocal tract. This study investigates changes in several acoustic variables for a transfeminine singer before and after undergoing surgery.

4.3 Methods

4.3.1 Data Processing

I used the Montreal Forced Aligner to automatically segment the speech recordings (McAuliffe et al. 2017). For the singing data, I used a singing-oriented forced aligner pretrained on English data.[2] For both speech and singing forced alignment, I used a dictionary for a US variety of English, as perceptually I deemed this the closest available option to Maria's variety of English. I systematically checked all forced aligner output boundaries, and manually corrected any errors.

I used a Praat script based on that of Vicenik (2009) to extract formants and H1-based voice quality measurements from all monophthongal vowel tokens, of which there were 1305 in the speech data and 3237 in the singing data. I calculated centre of gravity of all /s/ tokens, of which there were 142 in the speech data and 438 in the singing data, using a Praat script adapted from DiCanio (2013), which uses the methods described in Forrest et al. (1988). For formants and H1-based voice quality measures, measurements were taken at the midpoint of each token. Centre of gravity of /s/ was calculated over the middle 80% of each token. For the long term average spectra, I first normalised and manually end-pointed the recordings to remove silence from the start and end. For the singing data, I concatenated the recordings of individual songs into one sound file for each recording session. Mel-spectra were calculated over each sound file using Librosa's (McFee et al. 2023) melspectrogram function, with an FFT window of 2048 samples, hop size of 512, and a Hann window. 80 Mel-frequency bands between 0 and 8kHz were analysed. Long term average spectra were calculated by averaging the power spectra over time before computing dB relative to peak power.

4.3.2 Analysis

To explore the effects of surgery on the formant and H1-related voice quality measurements, I fit linear mixed effects regression (lmer) models to the data using using lme4 (Bates et al. 2015) in R (R Core Team 2019), for each dependent variable (i.e. F1, F2, F3, H1-H2, H1-A1, H1-A2, H1-A3). Speech models had recording date as a predictor and vowel quality as a random intercept. For singing models, song was also given as a random intercept. To explore whether the centre of gravity of /s/ changed in the singing data over the three recording sessions, I fit a linear mixed effects regression model, with centre of gravity as the dependent variable, recording date as the predictor and song as a random intercept. For the speech data,

[2] https://github.com/qiuqiao/SOFA/discussions/4

I fit a linear regression model, with centre of gravity as the dependent variable, and recording date as the predictor. In all models, recording date was given as a categorical variable, to account for the possibility of measurements not following a single trajectory across the three sessions. For formants and centre of gravity, frequencies were analysed in Hz. Since phonemic material was matched across recording sessions, there was no need to control for the position of the token in the word, utterance or sung phrase. For each of the models already described, an ANOVA was used to compare it with a null model which did not specify recording date as a predictor, to establish whether recording date had a significant effect on the dependent variables. For dependent variables for which the ANOVA yielded significant results, pairwise t-tests with Bonferroni correction were used to compare differences in measurement values between pairs of recording sessions, in order to establish where the significant differences lie. Measurements from singing and speech are not directly compared with each other, because the phonemic content is not matched across modalities in the data, and there was insufficient data to control for this and still make a meaningful comparison. For all statistical tests, the standard alpha level of 0.05 is adopted. Long term average spectra were not compared using inferential statistics (though they are of course correlated with H1-related measurements, which are analysed using inferential statistics), rather, I descriptively analysed the plots of the long term average spectra from each of the three recording sessions.

4.4 Results

Table 4.1 shows the results of the inferential statistics for formant, H1-related voice quality and centre of gravity of /s/ measurements.

4.4.1 Formants

Figure 4.1 shows the results of the formant measurements across recording sessions. Since the data was phonemically matched across recording sessions, vowel phonemes are not analysed individually here. The ANOVAs indicated that the effect of recording session was significant for F3 in speech, and for F1 and F3 in singing ($p < .001$ for each).

In singing, mean F1 shows an increase between each consecutive recording session, and is significantly higher in the 7 months post-surgery recording than the pre-surgery recording. In speech, F3 is significantly lower in the 3 months post-surgery recording compared with the other two recordings. In singing, F3 is significantly lower in the post-surgery compared with the pre-surgery recordings, and is significantly higher in the 7 months post-surgery recording compared with the 3 months post-surgery recording.

4.4.2 Voice Quality

H1-related Measures

Figure 4.2 shows the results for the H1-related voice quality measurements. The ANOVAs indicated that the effect of recording session was significant for all four H1-related measures. For speech, H1-H2, H1-A2 and H1-A3 all follow the same pattern of falling to a significantly lower value in the 3 months post-surgery recording compared with the pre-surgery recording, and then rising again in the 7 months post-surgery recording, to a level between that of

Table 4.1: *Results of the inferential statistics run on the acoustic measurements taken from the speech and singing data, where 'pre' corresponds to the pre-surgery recording session, 'post1' to the 3 months post-surgery session and 'post2' to the 7 months post-surgery session.*

		Effect of Date (p)	Bonferroni-corrected pairwise t-tests (p)		
			Pre-Post1	Pre-Post2	Post1-Post2
F1	Speech	0.05202	1.0000	0.1964	0.0654
	Singing	**0.000148**	0.1099	**0.0001**	0.0848
F2	Speech	0.1147	1.0000	0.4312	0.1325
	Singing	0.07888	1.0000	0.0830	0.2902
F3	Speech	**<0.0001**	**0.0006**	1.0000	**0.0001**
	Singing	**<0.0001**	**<0.0001**	**<0.0001**	**0.0004**
H1-H2	Speech	**<0.0001**	**<0.0001**	**0.0050**	**<0.0001**
	Singing	**<0.0001**	**<0.0001**	**<0.0001**	**0.0039**
H1-A1	Speech	**<0.0001**	**<0.0001**	0.3798	**<0.0001**
	Singing	**<0.0001**	**<0.0001**	**<0.0001**	1.0000
H1-A2	Speech	**<0.0001**	**<0.0001**	**0.0010**	**<0.0001**
	Singing	**<0.0001**	**<0.0001**	**<0.0001**	**0.0089**
H1-A3	Speech	**<0.0001**	**<0.0001**	**<0.0001**	**0.0018**
	Singing	**<0.0001**	**<0.0001**	**<0.0001**	1.0000
CoG /s/	Speech	**0.04337**	0.0961	1.0000	0.0830
	Singing	**0.03392**	0.2295	**0.0394**	1.0000

the pre-surgery and 3 months post-surgery recordings. H1-A1 in speech follows a similar pattern, except with the measurements from the 7 months post-surgery recording returning to the same level as the pre-surgery recording.

For singing, all voice quality measures are significantly lower in the post-surgery recordings compared with the pre-surgery recording. For H1-H2 and H1-A2, the 7 months post-surgery measurements are significantly higher than the 3 months post-surgery measurements, though still significantly lower than the pre-surgery values. For H1-A1 and H1-A3, the values in the 7 months post-surgery recording remain at the same level as those for the 3 months-post surgery recording. Since H1-A2 and H1-A3 follow the same pattern as either H1-H2 or H1-A1 for speech and singing, future discussion of H1-related measurements in this Study only focuses on H1-H2 and H1-A1.

4.4.3 Long-Term Average Spectra

Figure 4.3 shows the long term average spectra for the speech and singing recordings in the three different recording sessions. For the speech data, there is little difference between the three recordings. A slightly shallower spectral slope is visible for the 3 months post-surgery recording compared with the pre-surgery recording. The spectral slope of the 7 months post-surgery speech recording lies between those of the pre-surgery and 3 months post-surgery recordings.

In singing, there is a larger difference in spectral slope, with both post-surgery recordings

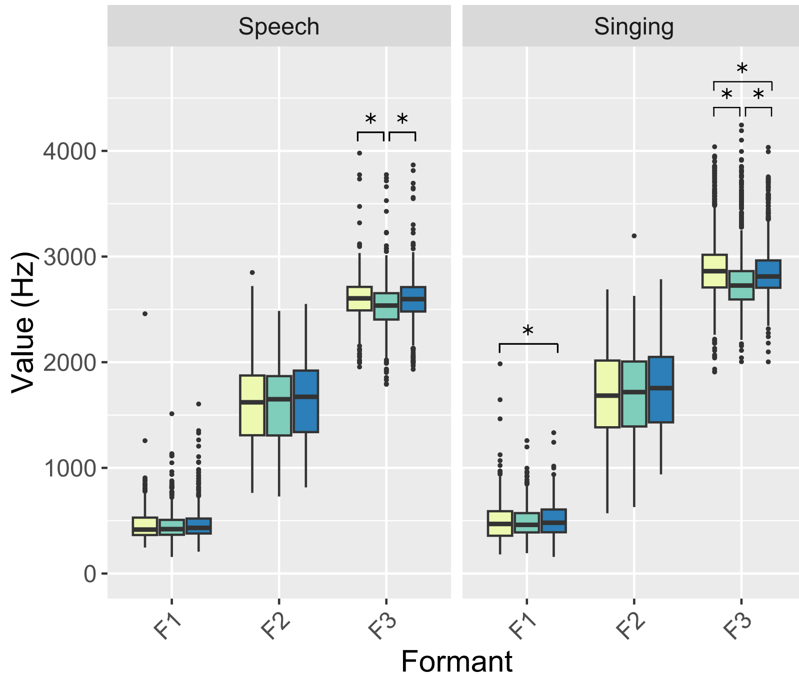

Figure 4.1: *Formant measurements across recording sessions for speech and singing data, for all monophthongs. Statistically significant pairwise comparisons are indicated with an asterisk.*

having a shallower slope than the pre-surgery recording, and little difference between the two post-surgery recordings.

4.4.4 Centre of Gravity of /s/

Figure 4.4 shows the results for centre of gravity of /s/ across the three recording sessions. Recording session was a significant predictor of centre of gravity in speech and singing ($p < .05$ for each). Centre of gravity values were overall high across all conditions. In speech, no pairwise comparisons between recording sessions were significant. In singing, mean values increased across consecutive recording sessions, and values were significantly higher 7 months post-surgery compared with pre-surgery.

4.5 Discussion

4.5.1 Formants

Recall that rhinoplasty and genioplasty were hypothesised to have an effect on Maria's formants, through physically altering the vocal tract. However, in addition to being potentially influenced by an altered vocal tract shape, formants can be changed with conscious or subconscious effort from the speaker, and altering formant frequencies is often a focus of voice and communication therapy for trans individuals, as higher formants are commonly associated with feminine voices (e.g. Carew et al. 2007). Since the significant raising of F1 over time

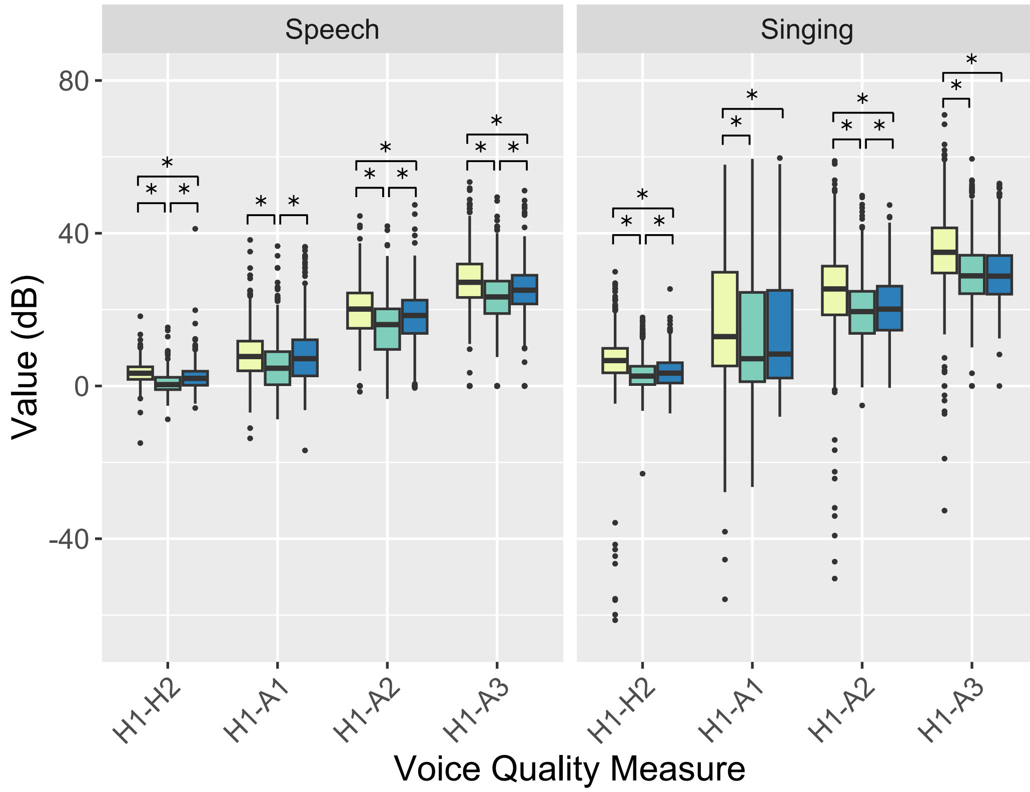

Figure 4.2: *H1-related voice quality measurements across recording sessions for speech and singing data, for all monophthongs. Statistically significant pairwise comparisons are indicated with an asterisk.*

is gradual and found only in singing, it is likely a result of Maria making a (sub)conscious effort towards more typically feminine vocal features in her singing. For both singing and speech, F3 was significantly lower in the 3 months post-surgery compared with the other two sessions. By 7 months post-surgery, F3 had returned to pre-surgery levels in speech, while in singing it had risen to somewhere between the pre-surgery and 3 months post-surgery levels. These results are consistent with a hypothesis that the surgery directly affected F3 for Maria, through altering her vocal tract. The raising of F3 again by 7 months post-surgery may be caused by Maria learning, either consciously or subconsciously, to actively compensate for this effect. This may have been done because the change caused by the surgery was not in the desired direction, since lower formant values are often associated with masculine voices in English (e.g. Avery & Liss 1996). This active compensation for surgical effects has been found for other facial surgeries (in presumably cisgender participants) (Bowers et al. 1985, Lee et al. 2002).

An alternative explanation for this pattern of results is that they may be due to Maria simply employing vocal feminisation practices relating to formant raising differently in the different recording sessions.

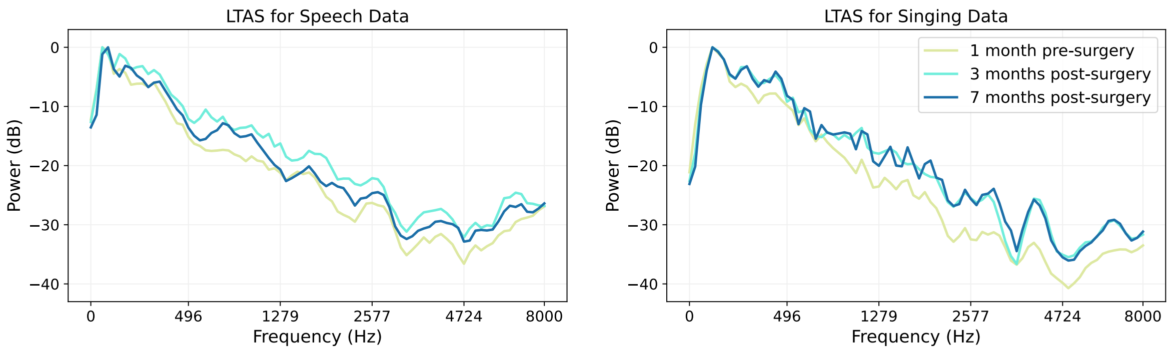

Figure 4.3: *Long-term average spectra (LTAS) for speech and singing data recorded in three different sessions.*

4.5.2 Voice Quality

Prior research shows that poor surgical technique for chondrolaryngoplasty can have an effect on voice quality (Nuyen et al. 2023). In particular, the procedure can cause hoarseness, a perceptual quality which Narasimhan & Vishal (2017) find to correlate with high H1-H2 and H1-A1. In the present study, however, H1-H2 and H1-A1 show a significant drop in the 3 months post-surgery recording. This indicates that Maria did not experience hoarseness as a result of the surgery. It seems unlikely, then, that the observed differences across recording sessions are due to physical changes caused by the surgery. Instead, they may be due to artistic decisions made by Maria on how to perform the material on different recording dates. Specifically, she may have opted for a less breathy reading of the Rainbow Passage in the 3 months post-surgery recording session and a more breathy performance of the songs in the pre-surgery recording session compared with the other recording sessions.

Alternatively, H1-H2 has also been found to correlate with vocal fatigue, with higher values corresponding with more fatigued voices (e.g. Narasimhan & Soumya 2020). Maria performed far more material before the analysed recordings in the pre-surgery recording session, so it is also possible the differing H1-H2 measures across sessions are due to differing levels of vocal fatigue.

The long term average spectra show shallower spectral slopes in the post-surgery compared with the pre-surgery recordings for singing and speech. Previous research finds women's voices tend to have a shallower spectral slope than men's (Mendoza et al. 1996). This is attributed to greater aspiration noise in the higher frequencies, and thus the results may be due to a (sub)conscious effort from Maria to shift her voice quality. Alternatively, this too may be due to artistic decisions or vocal fatigue, as discussed for the H1-related measures above.

4.5.3 Centre of Gravity of /s/

Overall, all centre of gravity of /s/ measurements are high, indicating that Maria was already adopting a very fronted production of /s/ (associated with femininity) before the surgery. For singing, measurements of centre of gravity of /s/ are significantly higher in the post-surgery recordings than the pre-surgery recording. These results seem indicative of Maria making a (sub)conscious effort to further feminise her production of /s/ in singing over time.

The results for speech are less clear, with a significant effect of recording session but no significant differences between pairs of recordings. Descriptively, there is a fall in centre of

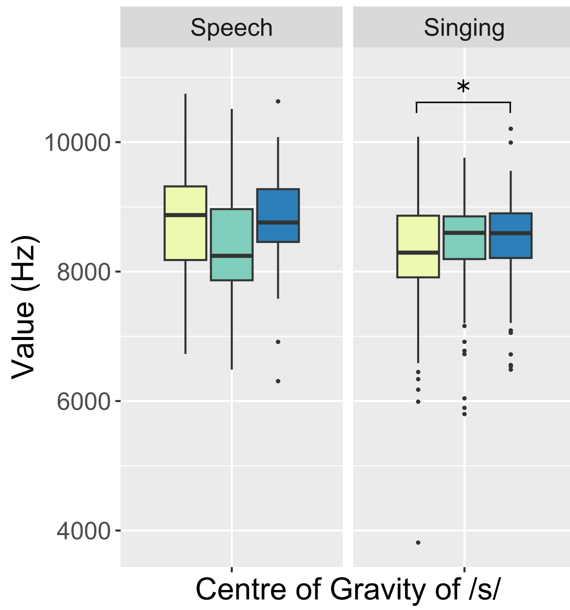

Figure 4.4: *Centre of gravity of /s/ measurements across recording sessions for speech and singing data. Statistically significant pairwise comparisons are indicated with an asterisk.*

gravity of /s/ between the pre-surgery and 3 months post-surgery recording, and then a rise to pre-surgery levels again by the 7 months post-surgery recording. Table 4.1 shows that the comparison between the pre-surgery and 3 months post-surgery, and the comparison between the 3 months post-surgery and 7 months post-surgery are both approaching significance with $p<0.1$. The comparison between the pre-surgery and 7 months post-surgery centre of gravity of /s/ values has $p=1$, on the other hand. It is therefore likely that the lack of significance for the two results with $p<0.1$ is due to a lack of power in this case, since there were far fewer /s/ tokens in the speech recordings compared with the singing recordings. However, further data would be needed to draw firm conclusions on these results.

4.5.4 General Discussion of Study 1 Results

In general, several measures taken from the speech data follow a pattern whereby the 3 months post-surgery recording shows a significant difference from the pre-surgery recording, and then the 7 months post-surgery recording shows either a return to pre-surgery levels, or to an intermediate level between the pre-surgery and 3 months post-surgery level.

Prior research investigating the impact of other surgical procedures on speech have sometimes found similar patterns of results. Bowers et al. (1985) investigated the effect of orthognathic surgery (corrective jaw surgery) on speech, through analysis of formant patterns for 3 vowels in 5 people at 4 different recording sessions: one before placement of orthodontic appliances, one before orthognathic surgery, one after orthognathic surgery and one after the removal of orthodontic appliances. All participants were judged to have perceptually normal speech both preoperatively and postoperatively by their respective orthodontists. They found that most participants showed a significant change in F2 for /i/ in the session following the orthognathic surgery, and then a return to pre-surgery levels in the session after the removal of orthodontic appliances. In this case, the authors concluded that the surgery directly impacted the participants' speech through changing the shape of the vocal tract, altering their F2, and that over time participants learned to alter their articulation, such that their acoustic values returned to their pre-surgery levels.

It is therefore possible that the pattern of results in the present study, for formants and voice quality measures, are due to the facial feminisation surgery altering Maria's vocal tract in ways which cause the differences in acoustic measurements found at 3 months post-surgery. If this is the case, then the results seen at 7 months post-surgery, which show a return to the pre-surgery levels, or at least a significant movement in this direction, could be due to Maria learning to adapt her articulation patterns in order to effect the same acoustic output as she did before the surgery. Evidence in support of this hypothesis comes from the fact that the changes in F3 and H1-related voice quality measurements observed at 3 months post-surgery compared with the pre-surgery recording all indicated that the surgery caused a move away from values typically associated with femininity in English. It seems unlikely that Maria would consciously employ speech features typically associated with masculinity (though see Zimman 2017). It therefore seems more likely that the significantly different measurements observed in the 3 months post-surgery recording were due to the changes in her vocal tract having a direct effect on her speech, rather than Maria consciously altering her speech.

If this is the case, then the same results should be expected for the singing data. And indeed, F3, and most H1-related voice quality measures do follow the same pattern. On the other hand, F1 and centre of gravity of /s/ rise linearly over the three sessions. As higher values of these variables are typically associated with femininity in English, these results suggest that Maria is actively working to adapt her use of these features. The different pattern of results in the singing data compared with the speech data may be due to the fact that singing is an explicit method of performance: it is less casual than speech and used less frequently, and it involves a greater focus on producing sounds in a specific way.

This pattern of results whereby there is a significant change in acoustic measures immediately after intervention, and then a return to pre-intervention or an intermediate level after several months is also seen in studies investigating the outcomes of gender-affirming speech therapy. Gelfer & Tice (2013) follows changes in trans women's speech patterns before, immediately after and 15 months after undergoing 8 weeks of gender-affirming speech therapy. In the recording session immediately following the speech therapy, participants showed a significant rise in average F0 and formants F1, F2 and F3 compared with their pre-therapy recordings. In the recording session 15 months later, participants' measurements had fallen again to levels between those of their pre-therapy and immediately post-therapy recordings. Applying this to the present study's results suggests that Maria may have been actively altering her speech at 3 months post-surgery, and that over time her values returned to similar levels to her pre-surgery speech, in the same way as the participants in the study of Gelfer & Tice. Evidence in support of this hypothesis comes from the centre of gravity of /s/ results. Descriptively, the centre of gravity of /s/ results for speech show a pattern very similar to those of F3 and H1-related voice quality measures. There is a fall in centre of gravity of /s/ between the pre-surgery and 3 months post-surgery recording, and then a rise to pre-surgery levels again by the 7 months post-surgery recording. Centre of gravity of /s/ was not predicted to be affected by direct effects of the surgery on the vocal tract, and indeed the singing results show a different pattern, so it is unlikely that the surgery had an unanticipated effect on the vocal tract which altered /s/ production. This suggests that perhaps Maria did (sub)consciously alter her speech patterns to move away from values typically associated with femininity, and invites a re-analysis of the formant and voice quality results.

Clearly, for many of the results in this study, there are several plausible explanations. I do not provide any conclusions here, and instead use these results to highlight the fact that research around the effects of surgery on the voice is an area where disentangling social and physiological factors affecting the voice is extremely difficult, because the features expected to potentially change with surgery (in this case formants and voice quality) can also be changed through conscious or subconscious efforts from the speaker. In the case of facial feminisation surgery in particular, the acoustic variables predicted to change with surgery are also often used to index a speaker's gender, and thus may be being (sub)consciously being altered by the speaker, and they are also a target of gender-affirming speech therapy and other voice training.

There are some important limitations to note regarding this study. Firstly, this type of research would ideally be done in conjunction with surgeons, who can provide exact details of the surgical procedures undergone by the participant. Not having this data limits the usefulness of this research, since facial feminisation procedures are numerous and varied, and the results of the present study cannot be linked with specific changes to the vocal tract with any degree of certainty. Additionally, having only one participant of course also means that these results are not generalisable to the wider population: further research with more participants, done in conjunction with surgeons, is needed in this area, to draw generalisable conclusions and thereby provide trans and gender diverse people considering undergoing these procedures with detailed information regarding the potential vocal impact.

Secondly, some issues arise from the fact that the data was not collected for this purpose. For example, the data does not allow for analysis of changes to minimum and maximum F0, since we did not ask Maria to demonstrate the full extent of her range in the recordings. It is also important to note that there was far more singing data than speech data, again due to the recordings being collected for a singing voice conversion system, and this may have contributed to some of the differences in the patterns of results across the two modalities. Additionally, the amount of material that Maria performed in or before a given session was not controlled, and so, as previously mentioned, vocal fatigue may be a confounding factor for the voice quality measures taken. Future research should carefully consider the question of how best to collect data for this type of research.

This Study raises several further questions. One is whether the statistically significant acoustic changes in the Maria's voice observed here are perceptually salient. Study 2 in the next chapter therefore aims to identify whether other people can perceive differences in Maria's voice across sessions. A second question raised by this Study is around the cause of the significant acoustic differences observed in Maria's voice across recording sessions. Study 3 in chapter 6 provides an analysis of interview data with Maria from before and after surgery, which helps to provide further insight into this question.

Chapter 5

Study 2: Externally-Perceived Effects

5.1 Introduction

Although the acoustic analysis in Study 1 indicated there were significant differences in the voice between recording sessions, that does not necessarily mean that these differences are perceptually salient. For example, with sufficient data, measurements that are consistently different across recording sessions can be acoustically significantly different, but if the difference is extremely small, then it may not be perceivable by humans. Additionally, there is a large body of research demonstrating that acoustic differences often do not mirror differences that listeners perceive, in the task of speaker identification (e.g. Sambur 1975, Liu et al. 2024), as well as in psychoacoustic tests (e.g. Liberman et al. 1957, Fry et al. 1962, Pobloth & Kleijn 1999, Paliwal & Alsteris 2003, Quené 2007). Informal listening did not reveal any large differences between the recording sessions in Study 1, so a perceptual experiment was conducted, with 32 listeners, to determine whether other people perceived Maria's voice to be different before and after surgery. Stimuli were created by taking extracts of the recordings of Maria from Study 1, described in section 4.3.1. The experiment involved listening to pairs of clips, either from the same or from different recording sessions, and rating how similar the person's voice sounds in the two clips.

Information on the extent to which vocal changes caused by facial feminisation surgery are perceivable by others may be helpful to people who are considering undergoing facial feminisation surgery, because it provides a more tangible and interpretable measure of the changes to their voice than acoustic measures, for those who do not have a background in phonetics, acoustics, or other relevant discipline. Importantly, the purpose of this study is not for listeners to evaluate any changes that Maria's voice went through on parameters such as 'femininity', or 'pleasantness' because such value judgements are highly subjective, and because Maria's own feelings around her voice with respect to these parameters, and relationship with her changing voice, are more important. Rather, this study is aimed at establishing whether any changes are perceivable by people who do not know Maria, and, by extension, whether the statistically significant changes found in the acoustic analysis of Study 1 have any perceptual relevance.

5.2 Background

Study 2 provides a method for evaluating the perceptual relevance of the acoustic changes to the Maria's voice, as found in Study 1, without reliance on value judgements. Specifically, Study 2 consists of a voice quality similarity rating task. Section 5.2.1, provides a discussion of how previous studies have measured the effects of two potential causes of vocal changes: facial surgery and gender-affirming speech therapy, motivating the use of a voice quality similarity rating task. In section 5.2.2, the definitions of the terms 'voice quality' and 'timbre' are discussed, as well as previous research on voice quality ratings. Section 5.2.3 provides an overview of pertinent literature on perception of acoustic differences in the voice, detailing the relevance of the present study to this area.

5.2.1 Measuring Vocal Changes

Previous studies use a variety of methods to measure the vocal impact of facial surgeries, and of gender-affirming speech therapy. One type of facial surgery for which the vocal effects have been widely studied is orthognathic surgery, which targets irregularities in the jawbones. People requiring orthognathic surgery often have disordered speech as a result of their condition (Lathrop-Marshall et al. 2021), which orthognathic surgery aims to help correct. The surgery involves repositioning the jawbones, and can also involve removing, reshaping or adding bone. Since the procedure entails altering the position of active articulators, it is expected to have an impact on articulation, and altering the shape of the jawbones can also potentially impact on voice quality. Vilanova et al. (2023) conduct a systematic review and meta-analysis of studies investigating the impact of orthognathic surgery on the voice. The 18 papers described use some form of acoustic analysis, articulatory analysis and/or expert judgement (for example, by a speech therapist). Many of the papers find significant vocal changes following surgery, but there seems to be a lack of investigation of the impact of vocal changes on the patients, and the effect that the surgery has on patients' relationships with their voice, although the impact of the surgery on quality of life overall is widely researched (for an overview, see de Araujo et al. 2019).

Recent years have seen an increased interest in the effects of rhinoplasty on the voice (of presumably cisgender participants). Studies investigating the impact of rhinoplasty on speech typically explore acoustic changes, and, unlike those for orthognathic surgery, they also explore changes in self-perception of the voice (Foroughian et al. 2014, Guarro et al. 2019, Bakhshaee et al. 2021, Claros et al. 2021). For self-perception of the voice, these studies ask participants to complete the Voice Handicap Index (VHI) before and after surgery. The VHI is a questionnaire designed for use in speech therapy settings with clients who have a voice handicap (Jacobson et al. 1997). It is filled out by the speaker, and is used to measure their perception of their voice handicap, with the speaker asked to rate the extent to which they relate to statements such as 'my voice makes it difficult for people to hear me', 'I feel as though I have to strain to produce voice' and 'my voice problem upsets me'. The majority of participants in these studies are reported to be undergoing rhinoplasty for cosmetic reasons, and are not reported to have any voice disorders. This questionnaire is therefore ill-suited to the setting of rhinoplasty, as it does not reflect issues that may arise when a person's voice changes in a way that could not be described as disordered, and questions explicitly referring to a 'voice problem' can be difficult to answer for speakers who do not identify as having a voice problem. T'Sjoen et al. (2006) highlights issues with using the VHI in unsuitable contexts through their study of trans individuals' perceptions of their voice. In their study, trans individuals completed the VHI and answered an additional question on the gender they tended to be perceived as on the telephone. Trans participants' VHI scores indicated no handicap, but their responses to the additional question revealed they were often misgendered on the phone, demonstrating that the VHI does not adequately capture issues that people may experience with their voice outside of voice disorders.

The use of the VHI in an inappropriate context likely contributes in part to the conflicting results between studies investigating the vocal impact of rhinoplasty, with Bakhshaee et al. (2021) finding no difference in VHI scores before and after surgery, Guarro et al. (2019) finding evidence for a worsening of self-perception of the voice, and Claros et al. (2021) finding evidence for an improvement of self-perception of the voice. Claros et al. (2021) additionally collect responses to questions around the participants' perception of their voice outside of the medicalised context of the VHI questionnaire. These questions include 'How do you evaluate your overall voice status after surgery?' with the options being 'worse', 'unchanged' and 'better'. The results of these questions give a more valid indication of the speakers' self-perception of their voices, which in this case indicated an overall positive perception of the vocal changes induced by the surgery.

To summarise, while orthognathic surgery studies do a good job of assessing the perceptual and articulatory vocal impact using expert judgements, they do not include enough focus on changes in the patients' self-perception of their voice. On the other hand, rhinoplasty studies make an attempt at documenting changes in the patients' self-perception of their voice, but often do so using methods ill-suited to the context, and studies for neither type of surgery have assessed whether any vocal changes are perceivable by naive listeners. In contrast, the outcomes of gender-affirming speech therapy are generally measured using a more holistic approach. Schwarz et al. (2023) conduct a systematic review and meta-analysis of studies on speech therapy outcomes for trans women. The assessment measures used in these studies include acoustic measures, various types of assessments by speech therapists, and various types of self-perception questionnaires. At least two such questionnaires have been designed specifically for trans people, including the Transgender Self Evaluation Questionnaire (TSEQ) (see Davies & Goldberg 2006), and the Transgender Woman Voice Questionnaire (TWVQ) (Dacakis et al. 2013). These provide context-appropriate alternatives to the VHI, reducing the issues exemplified by the study of T'Sjoen et al. (2006); questionnaires designed specifically for the population they are used for, allow for much better evaluation of the impact of vocal interventions. In general, the use of acoustic analysis, input from voice professionals and self-perception measures allows these studies to capture an overall picture of the outcomes of the gender-affirming speech therapy.

Some older studies on the outcomes of gender-affirming speech therapy included perceptual evaluation methods which involved naive listeners (i.e. not speech therapists) giving judgements on voice recordings. Gelfer & Tice (2013), for example, recruited naive listeners to provide binary gender judgements for recordings of trans participants before and after therapy, and to rate the recordings for 'pleasantness', 'masculinity' and 'femininity'. However, such methods assign too much value to the subjective evaluations of external people, and insufficient value to the perception of the trans participant themselves. If the ratings by external people are unfavourable, this could also have a negative impact on the trans participant's own self-perception of their voice. Study 2 presents an alternative way to incorporate the perceptions of external people in the evaluation of voice interventions, in a way that does not involve value judgements, but still allows for information gathering on the perceptual relevance of any acoustic effects of the intervention, specifically: voice quality similarity ratings.

5.2.2 Voice Quality and Timbre

Voice quality is a term found in the research of a diverse set of (sub)fields, including forensic speech science, speech technology, speech therapy, phonetics and singing research. Despite its wide use, the definition of the term is disputed. It is traditionally taken to follow from a definition of sound quality or timbre (see e.g. Kreiman et al. 2003). The Acoustical Society of America (ASA) gives the standard definition of timbre as: 'that multidimensional attribute of auditory sensation which enables a listener to judge that two non-identical sounds, similarly presented and having the same loudness, pitch, spatial location, and duration, are dissimilar' (Acoustical Society of America 2013, reaffirmed 2024). This kind of definition is extremely broad, encompassing many acoustic and perceptual dimensions. Several researchers have aimed to define the constituent perceptual dimensions of voice quality for various purposes, including to aid speech therapists in analysing pathological voices (see e.g. Laver 1980, Wirz & Beck 1995, Hammarberg 2000), and to evaluate the voices of professional vocalists (e.g. Bele 2007). A broad range of dimensions are presented in these works, including sonority, ring, vocal fry, roughness, breathiness, hoarseness, strain, nasality and creak. There is partial overlap in the dimensions presented in different papers, but no definitive list. The term 'voice quality' is therefore extremely vague in its definition, which has attracted criticism in the past (for an overview, see Kreiman et al. 2003), but makes the term ideally suited to the purposes of the present study. Asking listeners to attend to a broad term like 'voice quality' or 'timbre', while ignoring other specified factors that are not of interest (like pitch and phonemic content), allows for inclusion of any differences between voice samples that listeners may perceive.

Previous research shows that, in voice quality perception tasks, participants vary in the dimensions they pay attention to when making judgements. In a study with 5 speech-language clinicians and 5 naive listeners, Kreiman et al. (1990) find that listener experience affects voice quality perception: naive listeners were reported to all use similar perceptual strategies, whereas the clinicians differed greatly in the acoustic dimensions they deemed most important for voice quality similarity rating. Furthermore, Kreiman & Gerratt (1998) assess inter-rater reliability when listeners are asked to rate samples for 'breathiness' and 'roughness', two constituent components of 'voice quality'. They find that, while listeners display high self-agreement, inter-rater agreement is low. This indicates that individuals were understanding the task correctly, and that their understanding of the terms 'breathiness' and 'roughness' was consistent throughout the task, but that different listeners do not share the same understanding of what these qualities sound like. Kreiman & Gerratt (1998) conclude that these issues arise from listeners comparing samples with idiosyncratic internal standards when making judgements, and that they are unable to split the notion of 'voice quality' into its constituent parts. They argue instead for methods which involve comparison of two samples, and which do not ask listeners to provide separate ratings for individual components of voice quality. For these reasons, Study 2 involves asking people to rate the similarity between pairs of recordings based simply on voice quality. In this work, 'voice quality' is taken to be equivalent to 'timbre of the voice'. In describing the task to participants, then, both the terms 'timbre' and 'voice quality' are used, in order to increase the likelihood that the task is understood as intended.

5.2.3 Perceptual Salience of Acoustic Differences

There is a large body of research highlighting the differences between acoustics and human perception of voice similarity. Naturally, the vast majority of this comes from investigating the differences between different people's voices, rather than differences between two recordings of a single speaker's voice. Often, these studies involve using multidimensional scaling analysis, whereby pairwise speaker similarity ratings are converted to a low dimensional space, where each point in the space represents a speaker, with speakers closer together having been rated as sounding more similar. The resulting dimensions are then correlated with acoustic features, such as F0, formants or voice quality measures, and those with higher correlation are taken as being the most perceptually salient to listeners when making similarity judgements.

Matsumoto et al. (1973) present an early study using multidimensional scaling analysis, in which the stimuli consisted of recordings of single vowel productions, where they found F0 and formants to be the strongest correlates to speaker similarity. These findings are replicated in the more recent study of Baumann & Belin (2010), who additionally ran a principle component analysis (PCA) on the acoustic features of the voice samples, finding that other features (e.g. jitter, shimmer) are in fact more informative than F0 in differentiating the speakers in their sample. This demonstrates that acoustic differences do not correspond to perceptual differences in a straightforward way. Nolan et al. (2011) conduct a similar study, but with short (roughly 3 second) clips of spontaneous speech, rather than isolated vowels. They, too, find F0 to be the most important acoustic feature in speaker similarity perception.

However, in their survey paper, Schweinberger et al. (2014) argue that these studies find F0 to be the most important cue to speaker similarity perception because they only analyse a limited number of acoustic features, and F0 happens to be the most perceptually salient of those analysed. Indeed, Kreiman et al. (1992) find the picture to be much more complex, with evidence for both individual listener differences in which acoustic features are most salient for a given pair of stimuli, and differences across stimuli pairs for specific listeners. This suggests a high level of subjectivity in perceptual salience of acoustic features for differentiating between speakers. It also suggests that humans use different perceptual strategies for different voices. Further evidence for this comes from a speaker identification study by Van Lancker et al. (1985), in which participants were asked to identify the speaker for a number of audio clips of celebrities speaking, some of which were played forwards, and others in reverse. Some acoustic cues to speaker identity are preserved in backwards speech, such as average F0, F0 range, and voice quality measures. Others are reversed, such as F0 contours, consecutive syllable durations and other temporal information. Participants were able to identify some voices accurately, but not others, when played backwards, again indicating that different acoustic cues are used for identifying different voices.

Naturally, the extent of the differences between voices also influences which acoustic features are being relied upon to differentiate them. Recent attention has turned to very similar voices, such as those of identical twins (San Segundo et al. 2016). However, few studies have directly aimed to contribute to knowledge on perceptual salience of acoustic differences using data from the same speaker. This is a secondary aim of Study 2.

5.2.4 Summary

Previous studies investigating vocal changes caused by facial surgery tend not to collect data on whether any vocal changes are perceivable by naive listeners. Studies investigating the effects of gender-affirming voice therapy occasionally do seek perceptual ratings of others, but these have historically involved ratings of 'femininity' and 'masculinity' as well as value judgements on 'pleasantness', which can be emotionally detrimental to the trans individuals undergoing speech therapy, if the ratings are unfavourable. In Study 2, a method for gathering information on the perceptual salience of acoustic changes to the voice caused by facial feminisation surgery, without relying on subjective value judgements is proposed, namely a voice quality similarity rating task. The term voice quality is adopted for its broad scope, allowing for the capture of any perceivable differences, outside of pitch, temporal information and phonemic content.

As a secondary aim, Study 2 intends to contribute to knowledge on the perceptual salience of acoustic differences, with similarity rating data from recordings of the same speaker before and after surgery, where prior research has largely focused on similarity ratings of different speakers. A multidimensional scaling analysis is used to help to identify which acoustic features listeners attend to when making voice quality similarity judgements.

5.3 Methods

5.3.1 Ethics

The listening experiment described in this chapter was not covered by the original ethics approval described in section 3. Separate ethical approval was granted by the Ethics Committee of the Department of Computer Science at the University of Sheffield to cover the collection of listener data. An amendment to the original ethics application with the Department of Music was submitted and approved, and Maria gave informed consent for the use of her recordings as stimuli in this experiment.

5.3.2 Participants

Thirty-two fluent speakers of English with no known hearing loss participated in this experiment. Participants were recruited via the researcher's personal networks. At the beginning of the experiment, they were asked to self-describe their gender, and to select their age bracket. Table 5.1 shows how the participants responded to these demographic questions.[1]

5.3.3 Materials

This experiment involved comparisons of pairs of singing stimuli, as well as pairs of speech stimuli. Stimuli for this experiment were created from the singing and speech recordings from Study 1, described in Chapter 3. Recall that the singing data consisted of 6 songs sung at

[1]This information is provided to give an indication of the demographics of the participants. Age and gender are not included as factors in the analysis, due to the small numbers of participants in particular categories.

	18-25	26-35	36-45	46-55
‘man’ or ‘male’	6	3	5	1
‘woman’ or ‘female’	4	9	3	0
‘non-binary’	0	1	0	0

Table 5.1: *Age and gender distribution of participants in Study 2.*

each of the three time points: 1 month before surgery, 3 months after surgery and 7 months after surgery. The songs were *Can You Feel the Love Tonight*, *I Have A Dream*, *Jingle Bells*, *Moon River*, *Silent Night*, and *You Are My Sunshine*. The speech data consisted of a reading of the Rainbow Passage at each of the three time points.

For the singing stimuli, I selected four separate lines of lyrics from each of the six songs. When selecting lines, I took the following criteria into account: separate lines should not contain an overlap in lyrics of more than three consecutive words,[2] and the final selected lines should display a diversity of melodic content. Additionally, one of these lines had to have been sung by Maria more than once in the data from one time point. I then extracted the rendition of these lines from the recordings from each time point. For the lyric in each song that was selected on the basis of it having been sung more than once, two instances were extracted from the recording from one of the three time points. This resulted in 78 singing extracts in total. Additionally, a further line and pair of recording sessions were selected at random for use in the practice trial. This was an extract from Moon River, and was taken from each of the pre-surgery and 7-months post-surgery recordings. All singing extracts were between 5 and 20 seconds in duration.

For the speech stimuli, I divided the Rainbow Passage into 13 segments, at phrase or sentence boundaries. I then extracted these segments from the recording from each time point. This resulted in 39 speech extracts in total. All speech extracts were between 5 and 15 seconds in duration.

Each participant heard 24 singing trials and 12 speech trials. Each trial involved a pair of recordings, which participants were asked to compare. The pair of recordings either came from different recording sessions, or from the same recording session.

For the singing trials, each participant heard 4 trials for each of the 6 songs. Within these 4 trials, the 4 extracts for that song, and each of the four time point combinations (pre-surgery with 3-months post-surgery, pre-surgery with 7-months post-surgery, 3-months post-surgery with 7-months post-surgery, and two recordings from the same time point) occurred once. The pairing of time point combination and extract was randomised by participant. For the speech trials, each participant heard 3 trials of each time point combination. The extracts used for each of these was randomised by participant. Extracts were not repeated across trials.

Pairs of extracts within a singing trial were matched for phonemic content, but, due to constraints in the data, pairs of extracts within a speech trial were not. This is a limitation of the study, which will be revisited in section 5.5.3.

[2]It was not possible to meet this criteria for all songs, as some involved too much repetition of lyrics. In these cases, lines were chosen simply so that there was not a complete overlap in lyrics between separate lines.

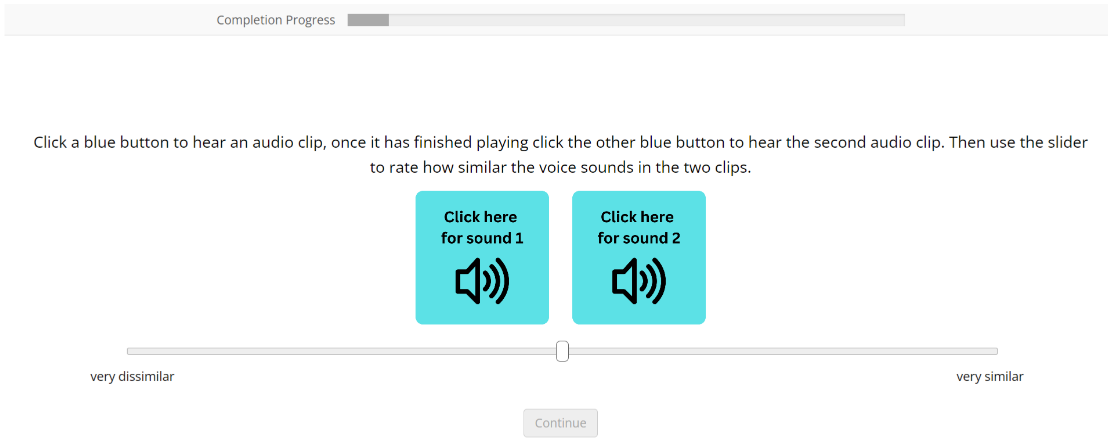

Figure 5.1: *Screenshot of a critical trial screen in Study 2.*

5.3.4 Procedure

Participants completed the experiment in a sound attenuating booth (IAC single walled), on a laptop, wearing a pair of Cooler Master CH321 headphones. The experiment was coded in jsPsych (de Leeuw et al. 2023) and ran on a web browser. Participants first answered demographic questions, which asked them to select their age bracket from a list, and to type their gender. They then saw an instructional screen, which explained that the first stage of the experiment would involve listening to pairs of recordings of a person singing, and rating on a scale how similar the voice sounds in each recording. Participants were encouraged to try to pay attention to the voice quality or timbre, and to ignore things like the words being said, or the key the song is sung in. They then proceeded to the practice trial, during which they were encouraged to adjust the volume to a comfortable level. They then proceeded to the critical trials for the singing part of the experiment. For the practice and critical trials, participants were presented with two blue buttons and were instructed to click on each of them to hear the audio clips, and to then use a slider to rate the similarity of the clips. Participants could listen to the two clips as many times as they wished. The left end of the slider was labelled 'very dissimilar' and the right end was labelled 'very similar'. After the slider had been moved, a continue button became clickable, allowing the participant to proceed to the next trial. A progress bar was visible throughout the experiment. Figure 5.1 is a screenshot of a critical trial screen. After completing 24 critical trials with singing stimuli, participants were instructed that they could optionally take a break, and that they could continue with the experiment whenever they were ready. Following this was another instructional screen, which explained that the next trials would involve clips of speech, rather than singing, and that the two clips would involve the speaker saying different words this time. Participants were encouraged to ignore this and to instead pay attention to the voice quality or timbre when responding. They finally proceeded to 12 critical trials with speech stimuli. These were visually exactly the same as the singing critical trials. The experiment lasted roughly 30 minutes, and listeners were given a £10 Amazon voucher in return for participating.

5.3.5 Analysis

Inferential Statistics

To evaluate whether participants rated samples taken from the same recording session as more similar than those taken from two different recording sessions, I ran a Kruskal-Wallis test in R (R Core Team 2019). The slider scores were values between 0 and 100, where 0 corresponded to the 'very dissimilar' end of the slider, and 100 to the 'very similar' end. Z-normalisation was applied per participant, and then min-max normalisation was applied overall (i.e. not per participant). These scores were then transformed to be in the range (0,1) and entered as the response variable. The predictor was the recording sessions being compared, with the categories being 'pre-surgery and 3-months post-surgery', 'pre-surgery and 7-months post-surgery', '3-months post-surgery and 7-months post-surgery', and 'same recording session'. Separate Kruskal-Wallis tests were run for the singing and the speech data. I then ran pairwise Wilcoxon rank sum tests with Bonferroni correction to identify differences between specific pairs of recording session comparisons. Additionally, for the speech and the singing data separately, a one-sample Wilcoxon signed-rank test was run on the responses from trials comparing two clips from the same recording session, to test whether participants responded significantly differently from chance for these trials. For all statistical tests, the standard alpha level of 0.05 is adopted.

Multidimensional Scaling Analysis

In order to give insight into which acoustic features participants were most salient to listeners when making similarity judgements, I employed multidimensional scaling analysis. This was conducted separately for singing and speech. Participants' normalised similarity ratings were averaged over stimuli, to provide one mean similarity score for each comparison of two different recording sessions. This was used to create a dissimilarity matrix, which was input to a classical multidimensional scaling (cmdscale) algorithm in R (R Core Team 2019). For this, 2 dimensions were specified, since only 3 recording sessions were being compared. I then ran acoustic analyses in Praat, on just the extracts from the recording sessions that were used as stimuli (for singing this represents a subset of the total recorded material, but for speech the entirety of the recorded material was used). Like in Study 1, formants F1, F2 and F3, as well as voice quality measurements H1-H2, H1-A1, H1-A2 and H1-A3 were extracted from the midpoint of all monophthong tokens using a Praat script adapted from Vicenik (2009). Centre of gravity of /s/ was extracted from the middle 80% of each /s/ token using a Praat script adapted from DiCanio (2013). In contrast to Study 1, here F0 and duration measurements were also taken, since these may have impacted listeners judgements of similarity. Specifically, vowel durations were recorded, and minimum, maximum and mean F0 for each stimulus was calculated using the autocorrelation method described by Boersma et al. (1993). The resulting 12 acoustic features were then averaged across tokens (if applicable) and extracts, to produce a mean for each recording session. Pearson's correlation coefficient was then calculated between each of the acoustic features and the two dimensions of the multidimensional scaling analysis. I performed this analysis procedure separately for the singing and speech data.

5.4 Results

Inferential and Descriptive Statistics

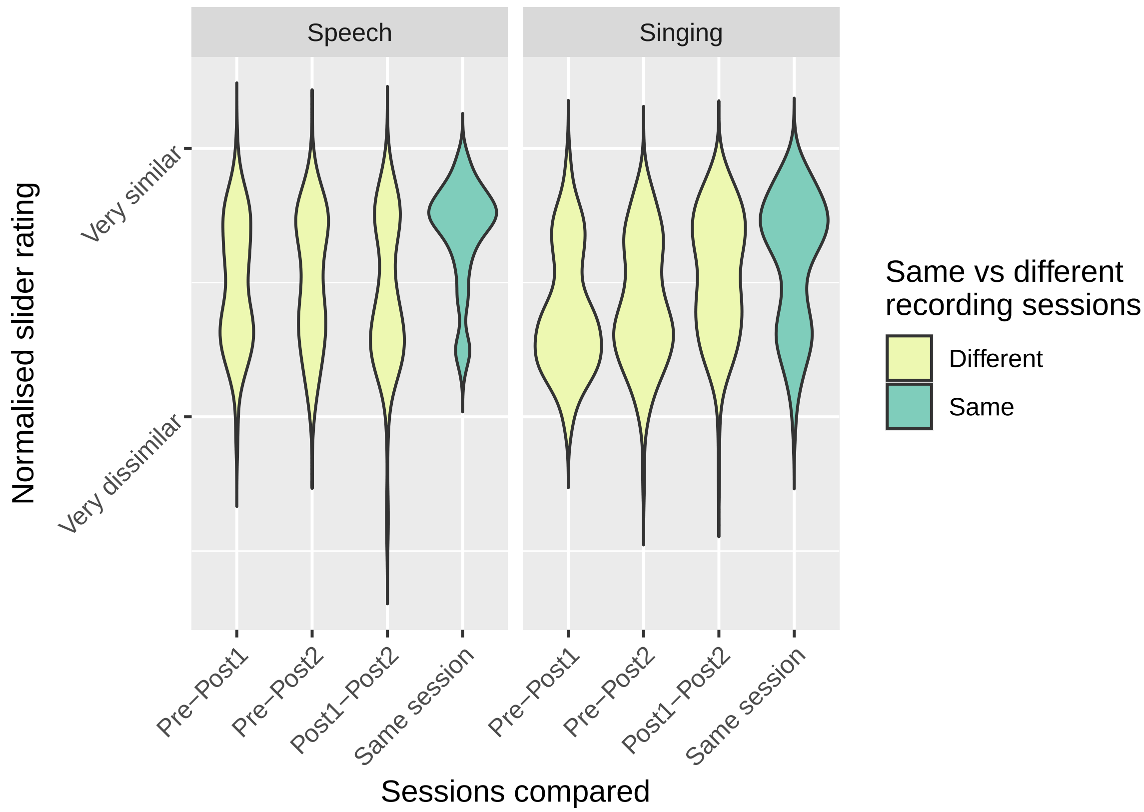

Figure 5.2: *Slider ratings for each of the pairs of recordings being compared in Study 2. 'Pre' refers to the pre-surgery recording session, 'post1' to the 3 months post-surgery recording session and 'post2' to the 7 months post-surgery recording session. Y-axis labels are located 2 standard deviations above and below the mean.*

Figure 5.2 shows a violin plot of the slider ratings given by participants in Study 2, according to the specific recording sessions being compared (where 'pre' corresponds to the pre-surgery recording session, 'post1' to the 3-months post-surgery session and 'post2' to the 7-months post-surgery session). Descriptively, almost all of the comparisons of 2 different sessions appear to have a bimodal distribution, with many ratings near 'very similar', as well as many ratings towards 'very dissimilar'. When participants were asked to compare two recordings from the same session, they mostly gave ratings close to the 'very similar' end of the scale.

For both singing and speech, the one-sample Wilcoxon signed-rank tests run on responses from trials comparing two clips from the same recording session yielded significant results ($p<0.001$ for each), indicating the mean slider score was significantly higher than 50 (i.e. closer to the 'very similar' end of the scale). The Kruskal-Wallis tests indicated that the recording sessions being compared was a significant predictor of slider score, for both singing and speech, with $p<0.001$ for each. Table 5.2 shows the results of the pairwise Wilcoxon

Table 5.2: *Pairwise comparisons using Wilcoxon rank sum test with Bonferroni correction, for singing.*

	Pre-Post1	Pre-Post2	Post1-Post2
Pre-Post2	0.710	-	-
Post1-Post2	**<0.001**	**<0.001**	-
Same session	**<0.001**	**<0.001**	0.140

Table 5.3: *Pairwise comparisons using Wilcoxon rank sum test with Bonferroni correction, for speech.*

	Pre-Post1	Pre-Post2	Post1-Post2
Pre-Post2	1	-	-
Post1-Post2	1	1	-
Same session	**<0.001**	**<0.001**	**<0.001**

rank sum tests for singing. Pre-Post2 & Pre-Post1 and Post1-Post2 & Same session were nonsignificant, and the remaining pairs were significiant, with $p<0.001$. In other words, for singing, participants' responses for trials comparing pre-surgery recordings with 3-month post-surgery recordings were similar to trials comparing pre-surgery recordings with 7-months post-surgery recordings. Their responses for trials comparing 3-months post-surgery with 7-months post-surgery recordings were similar to those comparing recordings from the same session. Table 5.3 shows the results for speech. All comparisons with the 'same session' category were significant, and all other comparisons were nonsignificant. This shows that, for speech trials comparing pairs of audio clips from different recording sessions, participants responded in a similar way, regardless of which recording sessions were being compared. They responded in a significantly different way, when the clips came from the same recording session.

Multidimensional Scaling Analysis

Figure 5.3 shows the results of the multidimensional scaling analysis for the singing recordings, with the three recording sessions plotted in a two-dimensional space based on the similarity ratings given by participants in Study 2. Here, the three recording sessions are relatively far apart, with the two post-surgery recordings slightly closer to each other than to the pre-surgery recording. Looking at the two dimensions individually, in dimension 1, the two post-surgery recordings are very close together, with the pre-surgery recording relatively far away. In dimension 2, the 7 months post-surgery recording and the 3 months post-surgery recording are relatively far apart, with the pre-surgery recording located in the middle of them.

Figure 5.4 shows the Pearson's correlation coefficients for each of the acoustic measures taken from the Study 2 singing stimuli with each of the two dimensions depicted in figure 5.3. Looking first at dimension 1, shown in red, the strongest positive correlations are with the voice quality measures H1-A3, H1-A2, H1-H2 and H1-A1, in that order. Strong negative correlations are found with the mean, minimum and maximum F0 measures, in that order. In dimension 2, shown in blue, most correlations are relatively weak, in contrast with dimension

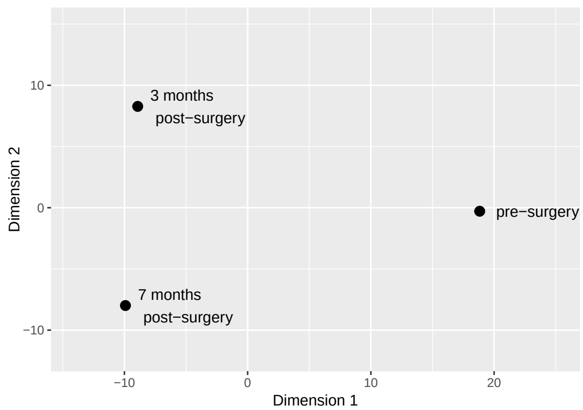

Figure 5.3: *The distribution of the three recording sessions in the two-dimensional space calculated in the multidimensional scaling analysis for the singing recordings.*

1. The strongest correlations with dimension 2 are centre of gravity of /s/ and F2, both of which are negatively correlated with dimension 2. Mean vowel duration and F1 were only relatively weakly correlated with the two dimensions. In general, most acoustic measures correlated very highly with one dimension or the other, which renders the analysis fairly uninformative. This result was observed because, as described in Study 1, many of the acoustic measures follow a similar pattern across the three recording sessions. However, this is not the case when looking at the acoustic measures for each song separately, and so multidimensional scaling analysis was then performed on the data from each song individually. For each song, an acoustic measure was considered to correlate 'highly' with an MDS dimension if it was the measure with the largest absolute value correlation coefficient, or if it had a correlation coefficient within 0.05 of the measure with the largest absolute value correlation coefficient. This resulted in a separate list for each song, of acoustic measures which correlated highly with a MDS dimension. Each of these measures had an absolute value correlation coefficient of at least 0.8.

Table 5.4 shows for each acoustic measure, the number of songs for which it correlates highly with a MDS dimension. No measure correlates highly with a dimension for all 6 songs. H1-H2, H1-A1, H1-A3, centre of gravity of /s/ and minimum F0 all correlate highly with a dimension for 5 songs. Formants and maximum and mean F0 all correlate highly with a dimension for 3 songs or fewer.

Figure 5.5 shows the results of the multidimensional scaling analysis for the speech recordings. In this case, the three recording sessions are roughly equidistant from each other in terms of Euclidean distance. Examining the two dimensions separately, in dimension 1, the pre-surgery and 7 months post-surgery recordings are very close together, and the 3 months post-surgery recording is a considerable distance from them. In dimension 2, values decrease consistently with linear time. That is, the pre-surgery recording holds the highest value for dimension 2, followed by the 3 months post-surgery recording, and then the 7 months post-surgery recording holds the lowest value for dimension 2.

Figure 5.6 shows the Pearson's correlation coefficients for each of the acoustic measures with the two multidimensional scaling analysis dimensions for the speech recordings illustrated in figure 5.5. Like for singing, dimension 1 is strongly positively correlated with all

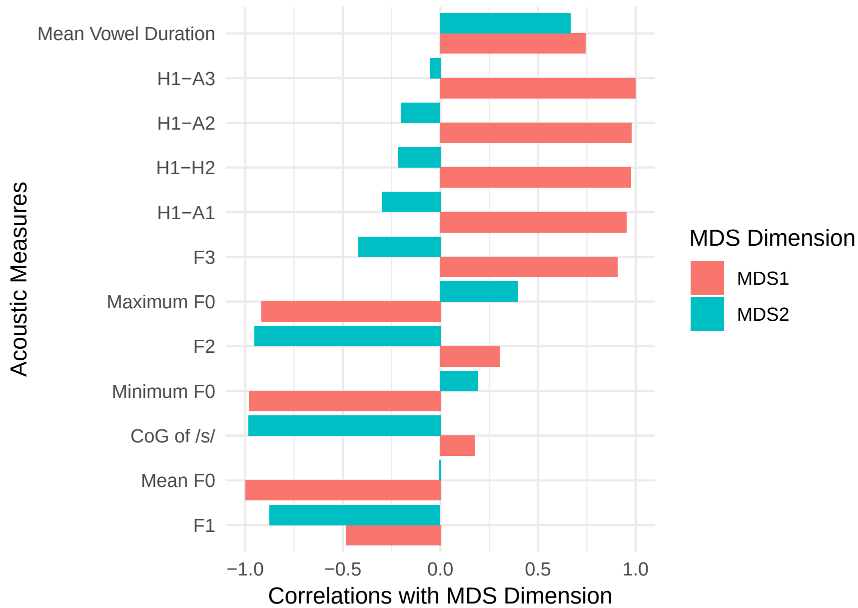

Figure 5.4: *Pearson's correlation coefficients of each of the acoustic features with each of the two dimensions in the multidimensional scaling analysis for the singing data in Study 2.*

voice quality measures, as well as with centre of gravity of /s/ and F3. Dimension 1 is also strongly negatively correlated with minimum F0. Again, dimension 2 correlates less strongly with the acoustic measures in general. The strongest correlations are the negative correlations with mean F0, F2 and F1, in that order. Maximum F0 correlates relatively weakly with the two dimensions.

5.5 Discussion

5.5.1 Inferential and Descriptive Statistics

The results of the one-sample Wilcoxon signed-rank tests on the results for 'same session' comparisons were significant. This can be viewed as a sanity check, indicating that participants were not simply responding at random, and generally judged the clips from the same session to be similar to each other.

Looking first at singing, and taking the inferential and descriptive statistics together, the results indicate that listeners overall judged the recordings from the same session to be very similar to each other. They also perceived the two post-surgery recordings to be very similar to one another, as their slider responses for the post1-post2 comparisons were clustered at the 'very similar' end of the scale, though to a slightly lesser extent than for the same session comparisons. On the other hand, though there are some responses for the pre-post1 and pre-post2 comparisons at the 'very similar' end of the scale, more are located towards the 'very dissimilar' end, indicating that participants perceived them to be more dissimilar from each other than clips from the same session or from the two different post-surgery sessions. Overall, these results indicate that participants perceived a difference between the pre-surgery

Table 5.4: *The songs for which each acoustic measure correlates highly with a MDS dimension.*

Acoustic Measure	Can You Feel the Love Tonight	I Have a Dream	Jingle Bells	Moon River	Silent Night	You Are My Sunshine	Count
F1	x		x	x			3
F2	x						1
F3		x		x		x	3
H1-H2	x	x	x	x	x		5
H1-A1	x	x	x	x	x		5
H1-A2	x			x	x	x	4
H1-A3	x	x	x		x	x	5
Duration		x		x	x	x	4
Min F0		x	x	x	x	x	5
Max F0	x	x	x				3
Mean F0					x		1
CoG /s/	x		x	x	x	x	5

recording and the post-surgery recordings for singing.

Turning now to speech, the results show, again, that the participants deemed recordings from the same session to be very similar to each other. They perceived recordings from all three different sessions to be noticeably different from each other. Ratings for each pair of different recording sessions were judged overall in a similar way. This indicates that, not only did the listeners perceive a difference between Maria's voice before and after surgery, they also perceived a difference in her voice 3 months after surgery compared with 7 months after surgery.

In general, the results for many of the categories in both singing and speech seem to have a bimodal distribution, which is likely due at least in part to the experimental design. Participants were required to move the slider in order to progress to the next trial. The starting position of the slider was in the centre of the scale, at 50, so although it was possible to give a score of 50 (by moving the slider away from the starting point and then back), this required extra effort and was therefore indirectly discouraged. The bimodal distributions also indicate that the acoustic differences between the recordings taken before and after surgery were only marginally salient. For all categories, there is a cluster of responses towards the 'very similar' end of the scale and another closer to the 'very dissimilar' end of the scale (after normalisation). Informal inspection of results for individual stimuli also indicates a high degree of variability across participants. Altogether, this indicates that the saliency of the differences differed across listeners and stimuli.

5.5.2 Multidimensional Scaling Analysis

In general, the multidimensional scaling analysis for the singing results was not very informative, since there were very strong correlations between most of the acoustic measures and

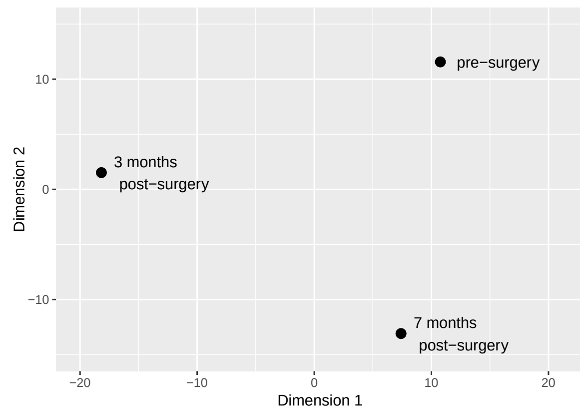

Figure 5.5: *The distribution of the three recording sessions in the two-dimensional space calculated in the multidimensional scaling analysis for the speech recordings.*

at least one dimension. This is because many of the acoustic measures patterned in a similar way across recording sessions, at least in the stimuli used for Study 2. This means that, it may be the case that participants were paying attention to both voice quality acoustic measures (H1-H2 etc.) and F0 measures (mean, minimum, maximum) when making their similarity judgements, or they may have been paying attention to just one of these measures, and the others happened to pattern with it. This kind of analysis is normally done with more than 3 categories (in this case the categories are recording sessions, in the wider speech literature the categories are typically speakers), and having more categories would make it less likely that all acoustic measures pattern in similar ways across the categories. Future work could take measurements at more than three time points in an individual's journey through facial feminisation surgery (or another voice-altering event) in order to gain more insightful results using this methodology. Comparing the results for individual songs, as in table 5.4, does give some additional insight. Three H1-related voice quality measures correlated highly with a MDS dimension for 5 songs, and the fourth correlated highly with a dimension in 4 songs. This indicates that voice quality, as measured by H1-related measurements, impacted on listeners' similarity judgements. Centre of gravity of /s/ also correlated highly with a dimension for 5 songs, suggesting this, too, was perceptually salient. Interestingly, no measure correlated highly with a dimension for all 6 songs. The likely explanation for this is that participants employ different perceptual strategies for judging similarity depending on the acoustic properties of the stimuli. This aligns with the results from speaker similarity literature, which indicate that people attend to different acoustic features depending on the voices involved (Van Lancker et al. 1985, Kreiman et al. 1992). The present results argue for the extension of this theory beyond the level of specific speakers and to the level of specific stimuli, and may be related to particular properties of the songs themselves. For example, Moon River is a slow song, and Maria's rendition of it involved lots of sustained vowels, whereas Jingle Bells is more upbeat, and this may have contributed to the difference in acoustic features that listeners seem to have attended to when making judgements across these songs. Future work could explore this possibility further.

All three formants correlated highly with a dimension for 3 songs or fewer. This indicates that formants were generally less salient to listeners than H1-related voice quality

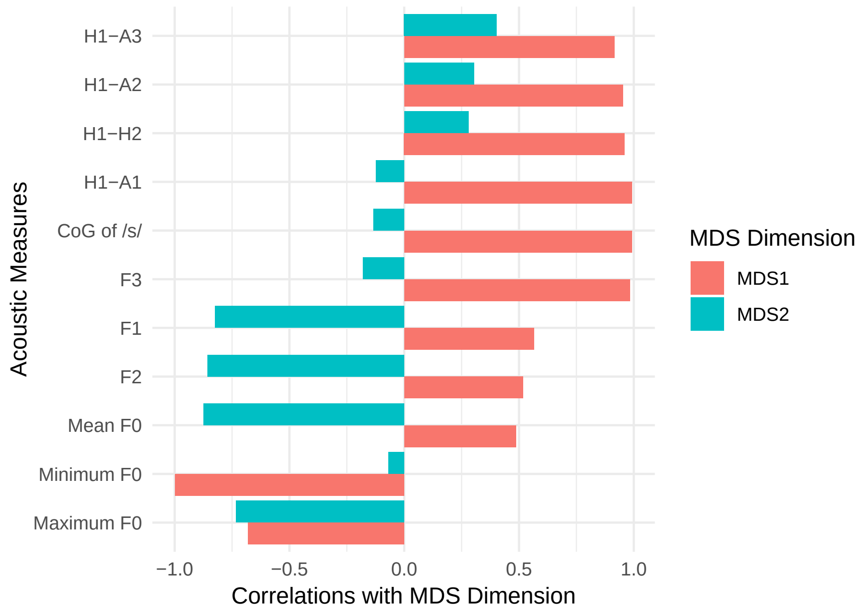

Figure 5.6: *Pearson's correlation coefficients of each of the acoustic features with each of the two dimensions in the multidimensional scaling analysis for the speech data in Study 2.*

measurements and centre of gravity of /s/.

For singing, minimum F0 correlated highly with a dimension for 5 songs, but maximum and mean F0 only correlated with a dimension for 3 songs and 1 song respectively. This indicates that participants may have been influenced by aspects of F0 in some cases, despite instructions to ignore pitch. Mean vowel duration also correlated highly with a dimension for 4 songs, indicating that participants may have been influenced by duration in some cases.

Turning to speech, again the results were not very informative, with voice quality measures (H1-H2 etc.), formants, centre of gravity of /s/ and mean and minimum F0 all showing strong correlations with one dimension. In this case, only maximum F0 can be interpreted to have had little influence on listeners judgements. This may have been due to the instruction to the participants to disregard pitch when making judgements, though the strong correlations of mean and minimum F0 prevent the concrete drawing of this conclusion.

It is worth reiterating that the recording conditions were kept constant throughout the three recording sessions, and so it is unlikely that participants' ratings were influenced by some auditory cue other than the voice (such as microphone characteristics or background noise).

5.5.3 General Discussion of Study 2 Results

Despite the limitations in the information provided by the multidimensional scaling analysis, this study nevertheless serves its purpose in establishing whether listeners were able to perceive a difference between Maria's voice in the three different recording sessions. In general, participants were able to do so, indicating that the statistically significant acoustic changes to Maria's voice established in Study 1 were perceptually noticeable. Though, it remains unclear whether all of the acoustic features measured in Study 1 were taken into account with participants' judgements, or whether only a subset were salient. It is also evident that the perceptual differences between Maria's voice in the three different recording sessions were only marginal, since there was a great deal of disagreement between listeners and stimuli on the similarity between pairs of voice samples.

Again, there are a number of limitations to this study which are important to note. Firstly, since this experiment was designed after Maria had undergone the surgery, the stimuli had to be extracted from the data that had already been collected for Study 1. For this reason, there are a number of potential confounding factors in the stimuli, that would normally have been controlled for in the experimental design process of a study investigating changes to a voice before and after an intervention like surgery or speech therapy. For example, no attempt was made to match the extracts being compared for speech rate, or for F0. Although voice quality and timbre are typically defined as not including pitch, F0 has been shown to have a large impact on similarity ratings (e.g. Matsumoto et al. 1973, Baumann & Belin 2010, Nolan et al. 2011), and there is also evidence that it interferes with timbre ratings (Melara & Marks 1990). Even though participants were instructed to avoid paying attention to pitch, it may still have subconsciously influenced their judgements. Speech rate (as approximated by mean vowel duration) seems to have had little influence on participants' judgements of the singing stimuli, but it may have had an influence on their judgements of speech stimuli. Additionally, although the phonemic content was matched across singing extracts being compared, it was not possible to do this for the speech extracts. It is known that certain phones provide more information on speaker identity than others (e.g. Amino et al. 2006), and so an experimental design in which speech extracts being compared were also matched for phonemic content would have provided a better comparison.

No information was collected on the L1 of the listeners in this study. Results on whether the L1 of the participants impacts speaker identification generally indicate that listeners are better at speaker identification in their L1 (e.g. Köster et al. 1997, Perrachione et al. 2009). It is possible that, for a voice quality similarity perception task like this, listener L1 would not have as much of an impact on judgements as in speaker identification tasks. This is because for voice quality similarity judgements between pairs of recordings from the same speaker, listeners are paying attention to more finegrained phonetic details than they are for speaker identity judgements. This is consistent with the results of San Segundo et al. (2016), which indicate that listener L1 has little impact on ratings of voice quality similarity for speakers with very similar voices (identical twins), when given short speech samples. Nevertheless, collecting information on the listeners' language backgrounds would have represented good practice, and would have allowed for verification that this finding still holds with longer samples, and with singing data. Additionally, there are a number of other factors that may have impacted listeners' judgements, including amount of musical training. Future studies could seek to also collect this information about participants, in order to clarify the extent of the impact they have.

Finally, it is worth reiterating that multidimensional scaling analysis is typically performed with a much greater number of categories. Performing it with just 3 here limits the conclusions that can be drawn from the analysis.

Overall, despite its limitations, this Study indicates that facial feminisation surgery caused some slight perceivable differences in Maria's voice. The question of whether these differences were caused directly by the surgery altering the vocal tract or by Maria herself making changes to her articulation patterns remains. Study 3 aims to bring some new insight into this, through qualitative analysis of interview data with Maria before and after undergoing surgery.

Chapter 6

Study 3: Self-Perceived Effects

6.1 Introduction

Studies 1 and 2 have shown that Maria's voice changed following facial feminisation surgery in a way that is both acoustically statistically significant and perceptually noticeable. However, the nature of the cause of these changes is still in question. This chapter describes a qualitative study that was carried out in order to gain insight into Maria's own perception of her changing voice, and her changing relationship with her voice, when undergoing facial feminisation surgery. This provides social context which is crucial for understanding the changes to Maria's voice.

Often, studies investigating the impact of facial surgeries on the voice do so through acoustic analysis of the voice before and after surgery, but can neglect to explore the impact of facial and vocal changes on the person themselves. Studies which do investigate changes in self-perception of the voice following surgery tend to do so with standardised questionnaires such as the Voice Handicap Index (VHI), which are often used in contexts for which they are unsuitable (see section 5.2.1 for a full discussion). Even when used in the context it is designed for, Kletzien et al. (2018) show that the VHI can fail to capture important details about a person's relationship with their voice. In their mixed methods study investigating the vocal impact of thyroidectomy, their interview data revealed that some patients experienced vocal impairment for at least 1 year following the surgery. However, this was not reflected in participants' VHI scores. Their study highlights the importance of conducting patient interviews when assessing the self-perceived vocal impact of surgery. Some studies investigate the impact of face-altering surgeries on quality of life more generally, but these results are never brought into the discussions around a person's vocal changes. Given the impact that social factors are known to have on the voice (e.g. Eckert 2012), clearly it is important to understand the emotional and social impact of surgery on the person undergoing surgery more generally, in order to fully understand reasons for vocal changes.

Study 3 presents a thematic analysis of interviews conducted with Maria at several points throughout her facial feminisation journey. This not only gives a potential insight into the underlying reasons behind Maria's vocal changes, but also provides a much more detailed insight into other aspects of her relationship with her voice.

6.2 Background

6.2.1 The Voice and Identity Revisited

As discussed in section 2.1, the voice is fundamentally linked to the speaker that produces it, through consisting of features which can reflect aspects of the speakers physiology and index aspects of their identity. The voice is also arguably unique to each individual, as indicated by the development of speaker recognition and verification models and their deployment in security contexts, for example.

Coming back to average F0, although it is known to be influenced by both physiological and social factors (see section 2.1 for a lengthier discussion), it provides an example of a feature which can be difficult for a speaker to consciously change, due to its relationship with physiology, as well as the fact that lifelong articulatory habits can be difficult alter. As a result, F0 and other physiologically influenced variables that are associated with different genders, can contribute heavily to vocal dysphoria in trans individuals. This is indicated by the evidence for a greater improvement in self-perception of the voice after hormonal treatment (with no corresponding voice training) for trans men compared with for trans women (Bultynck et al. 2017), since exogenous testosterone generally lowers mean F0 (Deuster et al. 2016, Hancock et al. 2017, Zimman 2017, Hodges-Simeon et al. 2021), but exogenous estrogen has little to no effect on F0 (Mészáros et al. 2005). On the other hand, phonetic variables such as the centre of gravity of /s/, which are used to index aspects of identity like gender and sexuality, relate to articulation rather than physiology and can therefore be manipulated by the speaker.

Consequently, such variables can be combined by speakers in different ways (in different contexts), in order to project various aspects of their identity. Zimman (2017) provides evidence for this, through their study investigating vocal changes in ten transmasculine individuals in their first few months of hormone therapy. Zimman conducted an ethnography and took recordings of readings of the Rainbow Passage (Fairbanks 1960) from each participant at monthly intervals over the course of a year. Centre of gravity of /s/ and mean F0 measurements are taken from the Rainbow Passage readings. These variables were chosen because the documented differences between people of different genders and sexualities in their centre of gravity of /s/ measurements are purely socially motivated, whereas mean F0 differences between genders are (at least partly) influenced by anatomy. Zimman finds that F0 generally decreases as time on testosterone increases, but that centre of gravity of /s/ measurements follow different trends for different participants. Several participants show a decrease in centre of gravity of /s/ measurements with time on testosterone, but one shows an increase; a participant who also described feeling more comfortable wearing clothes and using mannerisms more associated with femininity as time on testosterone increased and he became more consistently perceived as a man. In general, Zimman's results indicate that participants' low F0 provided them with the flexibility to use centre of gravity of /s/ to convey the nuances of their gender identity, gender presentation and sexuality. In this way, speakers are seen to contribute to the construction of their own vocal identity in complex and nuanced ways.

6.2.2 Vocal Loss and Vocal Dysphoria

Because the voice is linked to identity in so many ways, people can have complex feelings about, and relationships with, their voice. In the case of people who have lost or who are losing their ability to speak, evidence from speech synthesis indicates that their vocal identity is a core concern for them. Cave & Bloch (2021) conducted interviews with people living with motor neurone disease and their significant others, around their decision on whether or not to bank their voice. In this context, banked voices are used to create personalised speech synthesisers, with the aim that people who have lost their ability to speak can use a synthetic voice that sounds like them. They found that a key factor in the decision for many people was the idea of identity preservation. Those who decided to bank their voice did so because they felt it would help them preserve their identity when they could no longer use their natural voice. Those who decided not to bank their voice talked about how the resulting synthetic voice would still sound unnatural, or that it would be mimicking the dysarthric voice they now have, rather than their previous natural voice, indicating they did not believe it would adequately preserve their vocal identity. This demonstrates that many people are fond of their voice.

On the other hand, some people have a negative relationship with their voice. As mentioned earlier, the link between voice and identity, specifically the fact that certain phonetic features are perceptually associated with specific genders, means that for trans people, the voice can be a source of dysphoria. In general, there is a lack of research on the experience of vocal dysphoria in trans individuals using qualitative methods, but the few studies that do involve analysis of interviews about the voice from trans individuals have helped to paint a picture of the experience of vocal dysphoria. Holmberg et al. (2023), for example, provides an account of 15 trans people's experience of vocal dysphoria, through an analysis of interviews aimed at exploring motives for accessing gender-affirming speech therapy. Here, participants were found to experience dysphoria as both an internal feeling that the voice does not match with the self, and as a feeling caused or exacerbated by the (perceived or expressed) judgements of others. One participant described feeling that others believed they were lying about their gender identity because their voice did not align with listeners' expectations of how the voice of a person of their gender should sound. Vocal dysphoria was found to interact with other aspects of transition, with one transmasculine participant describing previously avoiding using feminine vocal features when they used to be frequently misperceived as a woman. However, after a year on testosterone, this participant was no longer misgendered, and so felt more comfortable incorporating some feminine vocal features, similar to the findings of Zimman (2017).

It should be noted that a wish to preserve the vocal identity in the face of the possibility of vocal loss and vocal dysphoria are not mutually exclusive. In this subsection I have talked about feelings that people have around their voice as a whole, but this is a simplification. The voice is of course multi-faceted, and so a person can have different feelings around different aspects of it: they could feel dysphoria around the pitch of their voice but still feel strongly connected to their regional accent, for example.

6.2.3 Management of Vocal Dysphoria

Vocal dysphoria can have a severe impact on the lives of trans individuals. It is linked with higher rates of depression (Junior & de Medeiros 2022), and it is also known to cause anxiety (e.g. Stewart et al. 2020, Holmberg et al. 2023), which can lead to avoidance of speaking situations, causing social isolation. For transfeminine people, vocal dysphoria is most commonly managed through gender-affirming speech therapy, or other forms of voice training. This is perceived as a very difficult process (at least by trans individuals who have yet to undergo any voice training), since it requires substantial time and effort, as well as potentially more time spent producing and listening to a voice which causes dysphoria (Holmberg et al. 2023). In the UK, gender-affirming speech therapy is (at the time of writing) available on the NHS, though this can be very difficult to access and waiting times are generally long. For people who do not have access to gender-affirming speech therapy with a qualified clinician, a number of other resources exist, such as YouTube videos and voice training guides made by other trans people.

Trans people manage their voice and experience vocal dysphoria in different ways in different contexts. Sport provides one example of a context in which binary conceptualisations of gender are still prevalent and upheld. Stewart et al. (2020) conduct interviews with transfeminine people about how their voice affects them in the sporting context. Many participants had undergone voice training and were able to speak in a voice that did not cause them dysphoria, but found that the physical exertion, high emotions and the need to project their voice in sport could make it difficult for them to maintain this voice. Participants who had strong relationships with their teammates and coaches reported that this caused them much more of an issue in during competition compared with training. Participants described feeling less pressure to maintain a feminine voice when talking to teammates and coaches compared with players from the opposition. This provides one example of how social context interacts with dysphoria and can affect the way transfeminine people use their voice. It also demonstrates the differing comfort levels that a person may experience when using their voice in different contexts, which can contribute to how they decide to use their voice.

Differing use of the voice in different contexts or with different people, or *style shifting*, is a commonly reported phenomenon outside of the context of trans people's speech. In his seminal work, *The Social Stratification of English in New York City*, Labov (1966) comments on differences observed in the frequency of usage of particular linguistic variables by individuals in different contexts. These include conversational compared with read speech, when a person is conversing with adults compared with when they are conversing with children, and informal compared with formal interview styles. In early sociolinguistic work, these differences were often thought to be the result of the differing amounts of attention paid to speech in these different contexts, and the primary focus of investigation was rather on the differences between macrosociological classes in speech, such as socioeconomic class, race and gender (see e.g. Eckert 2012). Later work saw style shift as worthy of investigation in its own right, with several linguists conducting case studies to uncover reasons why people style shift (e.g. Coupland 1980, Podesva 2007). Third wave sociolinguistics sees style shift as a practice which occurs as speakers actively use linguistic variables to construct styles according to their social goals and opinions of the interlocutors (see e.g. Eckert 2012).

Khoo & Ilbury (2024) provide an example of style shifting in the queer community more generally. Their sociolinguistic study follows the speech of two gay men across heteronormative and queer-friendly spaces in Singapore. They find that their participants, Rui and Kenni, style shift according to the levels of safety and comfort they feel in these two contexts, and how variables such as centre of gravity of /s/ carry negative connotations in the heteronormative settings, but are recontextualised and acquire positive meanings in queer-friendly environments.

6.2.4 Vocal Identity in Singing

Thus far, the discussion of voice and identity in this chapter has focused only on speech, but singers, too, craft a vocal identity for use in their work. Bette Midler and Tom Waits successfully sued Ford Motor Company and Frito-Lay respectively, for hiring people to imitate their vocal performances for advertising purposes (Stamets 1993, Cecchin 1992). This legal precedent in the US to protect artists when their vocal likeness is used without their permission provides an example of a field where the singing voice, like the spoken voice, is believed to be unique to individual.

In sociolinguistics, much attention has been given to the use of different accent features by vocalists when singing compared with speaking. Trudgill (1983) presents some early work in this area, which highlights the tendency for British artists to use General American features in their singing: features which are not generally present in their speech. This tendency is exemplified by English vocalist Adele, who speaks with a Cockney accent, but makes use of a number of General American features, such as a lack of the TRAP-BATH split, in her singing (Konert-Panek 2017). Scottish singer Nina Nesbitt provides a further example, with high rates of t-glottaling in her speech, but a tendency for [ɾ] realisations of /t/ in her pop songs (Papineau 2020). Other non-American singers have also been shown to adopt General American features in their work, including singers from New Zealand (Gibson 2019) and Ireland (Bencsik 2022). This phenomenon is commonly linked with Le Page's theory of linguistic behaviour, whereby "people create their linguistic systems [...] so as to resemble those of the groups with which from time to time they wish to identify" (Le Page 1986). Artists, then, are said to adopt linguistic features of American vocalists while singing, in order to align themselves with members of a tradition of commercially successful music, since the US has dominated the music industry in the last century. While this provides a plausible explanation for how this practice began, Gibson (2019) provides evidence that, at least for New Zealand-based artists, singing or rapping with General American features has become the default, and singing or rapping in one's native variety of English is more effortful. Artists who perform in their native variety of English (where this is not General American), then, are making a conscious choice to do so. Such a choice is more commonly taken for performances in certain genres, such as hip hop, where it is seen to mark authenticity, and resistance against the mainstream (Gibson 2019).

Naturally, for singers, the vocal identity also goes beyond accent features. It includes (but is not limited to) voice quality (encompassing the multitude of terms for aspects of voice quality used in singing, such as 'resonance', 'tone', 'colour', 'weight', 'timbre'), range, vibrato, projection, and the combinations of different settings of these variables in different musical contexts. A singer's vocal identity is developed and shaped over many years, with strong influence from others (Monks 2003, Sweet & Parker 2019). O'Bryan (2015) provides a qualitative analysis of interview data from a case study of an opera singer in training, Charlotte. Here, Charlotte's process of honing her vocal identity in singing was discussed. She describes being told upon beginning university-level voice training that her singing voice was not her authentic voice: that she had been mimicking opera singers rather than performing opera with her own voice. This again highlights the belief that in singing, as well as speech, the voice is unique to individuals. Following this, Charlotte had to both learn how to sing using her authentic voice, and then to learn to find the beauty in the way her authentic voice sounded. The impact on Charlotte of negative feedback was also discussed. Since, for singers, the body is the instrument, criticism of a singer's voice can feel like criticism of the singer themselves (O'Bryan 2015). Indeed for many singers, the voice is tied up with their sense of self, and as such some authors write about a different kind of 'vocal identity', using the term to encompass both the relationship that a person has with their voice, and the details of their identity as a singer (e.g. Monks 2003, Sweet & Parker 2019). In this way, the voice can be seen to both contribute to a singer's identity, and to reflect it.

The use of voice technology (or the choice to not use any) also forms a part of vocal identity construction for singers. Cher's 1998 song 'Believe' employs unconventional use of autotune, causing a notably 'unnatural' sound quality for the singing in the verses. It is often cited as the work that sparked mainstream use of this and similar voice-altering technologies in pop and other mainstream music. For her 2005 song 'Hide and Seek', Imogen Heap sang into a Digitech Vocalist Workstation harmoniser, which output vocoded harmonies at the pitches Heap played on a connected keyboard. In general, Heap's extensive use of technology in her work built her a musician identity so closely intertwined with technology that several news outlets and academics used a 'cyborg' metaphor in their descriptions of her (see e.g. Woloshyn 2009). More recently, Holly Herndon and Grimes provide examples of two singers whose voices are always distinctively processed in their released music. In this way, their characteristic technological processing of their voices forms a core part of their sung vocal identities presented to the public. Both Herndon and Grimes made the decision to release models of their voices, allowing the general public to use them in making music. In both cases, the output of these models is the artist's singing voice processed in their distinctive way, demonstrating the importance of technology to their sung vocal identities.

6.2.5 Authenticity and the Voice

Authenticity is a concept widely discussed in a number of fields, including music and linguistics. For a piece of music to be considered authentic, Coulter (2017) states it "must convey an impression of being real, raw, honest or original." Within linguistics, Bucholtz & Hall (2005) describe authenticity as the idea of which "sorts of language and language users count as 'genuine' for a given purpose." Authenticity is generally theorised as an attribute that a speaker or singer communicates or performs, and which listeners ascribe to a performance, performer or speaker, rather than something that the person or performance inherently has (e.g. Moore 2002, Bucholtz & Hall 2005). Judgements of authenticity, then, are subjective. Here I outline some previously documented beliefs around and judgements of authenticity in relation to the voice, from musicians, speakers and listeners, relevant to the present work.

Within singing, the notion of authenticity is often ascribed to a sense of 'unmediated expression' (e.g. Moore 2002). This could be because the distance between the singer's mental state and the produced song is perceived as being minimal, as Moore (2002) describes for Paul Weller's 'The Changingman', in which he employs a raspy voice quality, which Moore (2002) argues is suggestive of "a voice made raw from crying or shouting". The voice is also seen as being less authentic when mediated by technology. In the literature this is sometimes linked with the fact that the voice is produced by the body, and listeners perceive authenticity when they feel that what they are hearing is an unmediated reflection of the body (see e.g. Dickinson 2002). It is also linked with the fact that some members of the public see voice technology as a corrective tool or something used to conceal a lack of singing ability (see e.g. Coulter 2017).

Within speech, Holmberg et al. (2023) find that authenticity is also a concern for their participants. Trans participants who have yet to undergo gender affirming speech therapy express a desire to find a voice that feels authentic as well as in line with their gender identity, discussing times when modifying their voice has felt 'fake' or as though they were acting, rather than expressing themselves. This indicates that trans people search for authenticity in the voice on (at least) two levels: they wish for a voice that authentically expresses their gender identity, and one which is authentic to them as a person in a way that does not feel false or like an imitation.

6.2.6 Relationship Between the Visual Identity and Vocal Identity

Visual and vocal identity are closely linked, in both external- and self-perception. Since particular vocal traits are associated with particular groups, and aspects of visual appearance are also culturally associated with particular groups, when there is a perceived mismatch, this can raise issues. Revisiting O'Bryan's case study on opera student Charlotte, the relationship between Charlotte's identity as a singer and her body image also emerged as a theme (O'Bryan 2015). Charlotte talked about feeling the need to become fitter and to lose weight in order to physically be able to sing the repertoire she wished, and to be taken seriously in auditioning for particular roles. In this way, she perceived her visual identity as affecting both her ability to sing and her career prospects as a singer. She also talked about the idea, prevalent in opera, that body types should be congruous with voice types, with 'large' voices corresponding to larger bodies and 'small' voices to smaller bodies.

In linguistics, there is evidence that seeing a speaker's face impacts perception of the speaker's voice. Rubin (1992) conducted a study where two groups of American undergraduate students listened to the same recording of an L1 American English speaker giving a lecture. Each group was shown a photograph of a person and told that the person in the photograph was the lecturer. The first group saw a photograph of a white woman and the second group a photograph of an Asian woman. Following the lecture, participants completed a number of tasks, including rating the lecturer on a scale from "speaks with an American accent" to "speaks with a foreign accent". The group that saw the photograph of the Asian woman rated the lecturer significantly closer to the "speaks with a foreign accent" end of the scale than the group who saw the photograph of the white woman, despite the fact that both groups heard exactly the same audio. The listeners' perception of the speaker's voice was therefore influenced by what they believed to be the speaker's visual identity.

For trans people specifically, Holmberg et al. (2023) state that many of their participants discussed the impact of a mismatch between the voice and the visual identity. Participants described being read correctly as their gender identity when judgements were based purely on visual appearance, but that having a voice incongruous with their appearance meant that they were 'betrayed' when they started speaking. This also impacted their participation in daily life, since it had the effect of making some participants less likely to speak in certain situations. Similar sentiment was also expressed by the transfeminine participants in the study of Stewart et al. (2020).

6.2.7 Summary

The voice is known to be closely intertwined with aspects of a person's identity, and this is particularly true for singers. It is also commonly linked with ideologies such as authenticity, and in some contexts the voice is seen as needing to be congruent with a person's visual appearance. As a result, people can have complex feelings around their voice, including vocal dysphoria or a fear of vocal loss. People are also known to consciously or subconsciously use their voice to convey aspects of their identity, and to vary the ways in which they do this depending on the social context. As well as having the potential to affect aspects of a person's physical vocal tract, facial feminisation surgery alters a person's visual appearance. This can change the way people perceive a person, and therefore alters any social context that the person may find themselves in, as well as impacting the relationship between a person's visual and vocal identities. As of yet, however, no studies have investigated the impact of facial feminisation surgery on a person's feelings around, perception of and relationship with their own voice. Study 3 addresses this gap through a thematic analysis of recorded interviews with Maria before and after undergoing facial feminisation surgery.

6.3 Methods

6.3.1 Data Processing

The interviews were automatically transcribed using Whisper.[1] I then manually corrected the Whisper outputs. For the quotes used in this thesis, repetitions and instances of retracing with correction were removed, according to Maria's wishes and for ease of reading.

6.3.2 Analysis

I conducted a thematic analysis of the interview data, following Braun & Clarke (2006), using a bottom-up inductive approach, allowing the themes to be driven by the data, rather than by a particular research interest[2] or theory. I adopt a realist perspective, whereby the content of the interviews is seen to provide an insight into Maria's beliefs and experiences. Initial codes were made at the semantic level, identifying the surface-level meaning of what was said. I conducted a validation interview with Maria after themes had been defined and named (phase 5 of thematic analysis as described by Braun & Clarke 2006), following which, some details of specific themes were altered, in line with Maria's comments.

6.4 Results

Several themes were identified in the interview data. Only those that directly relate to Maria's voice are described here. In this section, I describe each theme in detail.

[1] https://huggingface.co/openai/whisper-medium accessed 28.8.24

[2] Since this thesis revolves around the voice, however, I only present themes specifically relating to this here.

6.4.1 Theme 1: Changes to the Voice and Vocal Tract Caused Directly by the Surgery

One of the themes identified was around the direct impact that the surgery had on Maria's voice and vocal tract. Where Study 1 of this thesis concerns the impact of the surgery on the acoustics of the voice, and Study 2 around the perceptions of others about the impact of the surgery on the voice, this theme details the changes to the voice that Maria herself could hear and feel, as well as how these fit with the expectations she had prior to undergoing the surgery.

In the first interview, Maria detailed the ways in which she expected her voice to change as a result of the surgery. Namely, she anticipated limited access to the higher notes in her range, either temporarily or permanently. She was also concerned about changes to her voice quality, noting specifically that the brightness of her voice might change. Overall, she expected the vocal changes caused by the surgery to be for the worse and expressed a great degree of uncertainty around what may happen to her voice. She primarily anticipated changes to her singing voice, stating: "I don't think my spoken voice is going to change dramatically enough to warrant further research into that right now."

In the second interview, she reflected on how it felt to sing after the surgery, noting that she largely retained control of her voice, which she attributes to the care of the surgical team.

> "I can feel that I have control of the vast majority of my voice. They were very careful."

She commented on how she felt a slight lack of control with regards to her voice quality, and in singing in the highest part of her range.

> "So in terms of resonance or in terms of like the feeling of the voice, it's been a bit like shooting in the dark [...] I would say maybe my highs right now are like, I don't feel as much in control of them, necessarily, but not in a way that I completely can't reach them."

This provides an example of a vocal effect of the surgery that is unlikely to be captured by other methods of investigating the vocal effects of surgery, yet could still have a profound impact on the life of a singer, and so is important to identify.

She then added that she had not yet explored the full extent of her singing voice post-surgery, so it felt too early to "make those very final assessments." And indeed, in interview three, once she had had time to fully test the extent of her voice post-surgery, she stated that she didn't feel any lasting effects of the surgery on her voice. In this case, then, Maria's negative expectations of how her voice would change with the surgery were not borne out.

Taken altogether, the findings of this theme both motivate the present research and highlight the need for further research on the impact of facial feminisation on the voice, with more participants, in order to provide people considering undergoing the surgery with a better picture of how their voices may change.

6.4.2 Theme 2: Vocal Dysphoria

The second theme identified centred around Maria's experience of dysphoria. In particular, she discussed her experience of vocal dysphoria and the ways in which the surgery affected her relationship with her voice.

In the interview prior to facial feminisation surgery, Maria described not feeling very much dysphoria around her voice. She had had some gender-affirming speech therapy sessions, since they were offered by her healthcare provider, but did not feel the need to put the techniques she had learned in these sessions in to practice very often when speaking:

> "I don't feel an intense amount of dysphoria. I feel some, but I don't feel an intense amount of dysphoria around the pitch of my voice, [...] so you're not hearing me apply all those practices that I've technically learned."

In singing, too, her voice did not generally cause her dysphoria. She describes feeling comfortable singing in the lower region of her pitch range, and feeling confident enough in her relationship with her voice that she did not believe that this would change:

> "I'm comfortable [...] singing in my lower range: it doesn't affect my gender identity or my relationship with my gender in any particular way. And I don't anticipate that's going to change."

Despite this, in the first interview she also acknowledged the potential for emotional changes to be brought on by the surgery:

> "This is a journey which, the immediate view that you can have on it is that it's going to be a journey of loss. But at the same time, I'm realigning so much, so intensely, to the gender presentation and the gender that I feel internally connected to. So, realistically, there's a very big psychological element there. We don't know what's going to open up. We don't know what's going to change in that way. And it is very much, in that sense, a journey of exploration and charting a new relationship between my gender, my gender expression and my voice."

In the second interview, she details a number of the emotional effects of the surgery, including feeling that there was now an incongruence between her face and other aspects of her person. She describes how, prior to the surgery, she saw herself making very small steps in her transition, so that all aspects of herself (e.g. face, voice) progressed at a similar pace. However, the facial feminisation surgery brought on a very significant and sudden change in her facial appearance, which caused her to feel that all other aspects of herself were lagging behind, in terms of the transition. She was particularly conscious of this feeling of incongruence between her face and her voice, wondering about how to use her bass-baritone voice in a way that makes sense with her feminine facial features, and talking about a newly-felt visceral connection with her vocal identity:

> "I look at myself in the mirror and I tell myself, well objectively this does not look like a man anymore, this is not a man's face, [...] but what do I do with this voice now that can go to a C2? What do I do with this voice that is effectively a bass-baritone? What do I do with this voice in a single setting? Especially as my transition continues, the external aligns, but the voice... do I push higher and higher? Do I push throughout my falsetto? And also thinking, well to a trained ear, that's always gonna sound like, that's a voice touched by testosterone."

She reports experiencing an onset of vocal dysphoria after the surgery, which she sees as a product of this perceived incongruence between the face and the voice. This described sensation of disconnect between the visual and the vocal identity mirrors the feelings of the trans participants in the studies of Stewart et al. (2020) and Holmberg et al. (2023). It also links to notions expressed by Charlotte in the study by O'Bryan (2015), about how it is believed within opera that particular singing voices should map to particular bodies. Though not an opera singer, as a singer in general, Maria also has an additional awareness around how she is perceived by an audience and not just by her interlocutors. As will be discussed in section 6.4.3, this leads to concern about how the vocal-visual mismatch could potentially negatively impact her singing career.

Another way that Maria frames her onset of vocal dysphoria is as a result of the reduction in facial dysphoria. She describes dysphoria as something that travels through the body, in a similar way that grief does, and that the facial feminisation surgery has caused some of her dysphoria to move away from her face and towards other aspects of her person, such as her voice.

> "Grief travels through the body, [...] those emotions, they move through you. And to me, dysphoria, to a big extent, is of a similar nature. So the dysphoria that's been concentrated in my face, a part of it has been sawn away, a part of it has been bossed and grinded down and grinded away, but a part of it has just shifted. And it's shifted to things like my shoulders, my voice, shifted to other parts of my body."

She clarifies in the validation interview that she does not believe that the vocal dysphoria was not present before the facial feminisation surgery, rather that the dysphoria she felt around other aspects of her person, such as her face, overpowered the vocal dysphoria:

> "There's only so much cognitive attention that you can dedicate to suffering, [...] I don't know if you can hate all of yourself at the same time."

Her description of dysphoria changing in intensity and moving around the body following surgery adds to what is known about dysphoria, the voice, and the wider impact of gender affirming healthcare on these. Her experience is consistent with prior results from Zimman (2017) and Holmberg et al. (2023), which show vocal dysphoria to interact with other aspects of transition. However, in their transmasculine participants, vocal dysphoria was found to ease as other aspects of their person began to convey their gender to interlocutors more clearly, and they then became more comfortable incorporating feminine vocal features into their speech. In the present Study, the opposite is the case. A potential reason contributing to this is that the range of voices which naive listeners perceive as male is perhaps broader than the range they might perceive as female.

She talked about reckoning with the feeling that her transition is hitting a plateau in certain ways, and the fact that there are aspects of herself that she will never be able to change, like the width of her shoulders, and her height. She described struggling to cope with this feeling, since a core theme of transition is to take control of your gender presentation.

> "There is a realization there which I've always had and it's nothing new, but it's very visceral for the first time now, that there are certain things that you will live

> with for the rest of your life, and you will have to find a way to make peace with. And it's a hard feeling to contend with sometimes, because so much of transition is about taking control."

Not experiencing negative consequences of the surgery on her voice was very unexpected for Maria. She spoke about how making the decision to undergo the surgery was extremely difficult, given the potential risk to her voice. She had decided to put her transition over her identity as a singer, and accept the possibility that she may have to forego her singing career. The fact that this outcome was never realised brought on unexpected emotions for her. In the second interview, Maria described feeling undeserving of the fact that she experienced no negative impacts of the surgery on her voice, given that she had been willing to give up her singing voice completely.

> "I was willing to sacrifice this part of myself that's been such a huge part of me, and now I'm potentially looking at continuing my creative career, where I make money out of it, where I'm successful through it. [...] There's almost a slight moment of how much do I deserve this? How much do I deserve the fact that I got no damage, given that I was willing to sacrifice this?"

In the validation interview, she describes how many of these emotional effects of the surgery have subsided over time, in particular the feelings of not being deserving. She also talks about how her recent experiences with other trans people helped reduce her dysphoria around her height and build, and therefore feel less troubled by the feeling that her transition had hit a plateau.

In this theme, Maria provides some valuable insight into the nature of (vocal) dysphoria. Naturally, trans people who do not feel vocal dysphoria are typically not included in qualitative studies exploring the phenomenon, and so experiences like Maria's are not often documented. Maria experienced a drastic change in her relationship with her voice following facial feminisation surgery, which she explicitly did not anticipate. This is not only an impact of the surgery on the voice which would not have been captured by acoustic or perceptual investigation, but it also begins to suggest that Maria may have begun making a conscious effort to change her voice after surgery, and thus contributes to our understanding of the acoustic results obtained in Study 1.

6.4.3 Theme 3: Vocal Goals and Ways to Change the Voice

Indeed, Maria also talked explicitly about her changing vocal goals, and various options available to her for pursuing these. In the first interview, she did not express any desire to change her voice, likely due to the fact that she did not feel an intense amount of vocal dysphoria at this stage. In the second interview, following the onset of her vocal dysphoria, she mentions several possibilities she is considering for changing the ways in which she uses her voice, particularly in singing. These include ceasing to use the lower end of her pitch range, so that she never sings below the lower bounds of an alto's range, and training herself to be able to sing higher notes. However, she seems uncertain as to whether these are things she definitely wants to pursue. In the third interview, Maria's vocal goals are much more concrete. She is working with a vocal coach and is focused on exploring new creative ways to use the tools that she has in ways that are more affirming. This includes experimenting more with her falsetto and mixed voice. She is no longer aiming for vocal feminisation, rather androgynisation, as she feels that that is a goal which is both affirming, and achievable and sustainable given her bass-baritone starting point:

> "There's a big vocal range where I can feel affirmed in myself, but it can also be sustainable. And that's for the most part been the androgynous space."

In the validation interview, she confirms that she is still happy with these goals and is actively working towards them, and she has begun to use her falsetto and mixed voice in performance.

In terms of her creative practice, she talks also about continuing to regard her older solo work, which has her bass-baritone vocals on it, in a positive light. However, she has concerns about what to produce next, as bass-baritone singing no longer feels authentic:

> "I'll release my debut EP which has my fairly bass-baritone vocals on it. I don't mind that, you know, I stand behind the work. But then after that, what do I create? How do I create something that still resonates?"

She looks to other popular trans artists, and describes a trend that she sees, whereby transfeminine artists with higher vocal ranges and more feminine appearances, tend to have a more traditional approach to artistry, where the person's visual identity and personality is used to promote their music. On the other hand, she sees transfeminine artists with lower vocal ranges tending to make more use of voice technology, and in many cases make use of a kind of depersonalisation, whereby the artist's authentic visual identity is not directly used in the promotion of the music. Maria is concerned by this, since she sees excessive use of voice technology as obscuring authenticity, and emotional connection with the audience.

> "[Voice technology] is something that I 1000% intend to include into my practice, and into my performance sets and into my music, but would I be hiding exclusively behind [it]? I guess even using the word 'hiding' kind of lets out the way that I think about it. But would I be using that 100% of the time or would I be using it at choice moments, to widen the creative range that I want to communicate with the listener? And I think for me it would probably be the latter [...] because that's probably more the kind of creative practice that I belong to I think."

Authenticity is seen to be a goal in several areas of Maria's life and work. She sides with previously documented beliefs that voice technology can serve to obscure authenticity, since it acts as an additional barrier between a singer's emotion and the sound a listener perceives (e.g. Coulter 2017). However, the lack of commercially successful transfeminine artists who have deeper vocal ranges and openly use them in their music, raises concerns for Maria about the commercial viability of her own creative work. She finds her pursuit of authenticity potentially at odds with success in the music world.

Throughout the interviews, she talks about different methods available to trans people more generally for changing the voice, and her experience of and attitudes towards these. Specifically, she discusses vocal feminisation surgery, gender-affirming speech therapy, and the vocal impact of hormones. In the first, second and validation interviews she consistently expresses a strongly negative attitude towards vocal feminisation surgery. She states that it has never been something she has considered, as it is "a complete gamble, and [she] would not expect to have any voice left," and because it is strongly discouraged by medical professionals for anyone who uses their voice professionally. The effects of vocal feminisation surgery are also largely permanent. In the validation interview, she mentions that she knows several people who have undergone the procedure, with negative results.

Maria talked about the four sessions of speech therapy she had prior to undergoing the surgery, where she was taught some techniques for feminising her speaking voice. She also had 2-3 sessions of singing-focused vocal coaching at that time, which she felt aligned with the speech therapy in many ways, and helped her to fully grasp the techniques she was learning in the speech therapy sessions. At the time, she did not actively put these techniques into practice, as the fact that she did not experience an intense amount of vocal dysphoria meant that the concept of 'feminising' her voice did not feel relevant to her. By the third interview, however, she had found herself relying on these techniques more and more, following the increase in vocal dysphoria after the facial feminisation surgery. In particular, she described having begun speaking with a lighter voice quality, and at a pitch that was somewhat higher, though still within a range that is comfortable to speak at for several hours. She describes these changes in her speaking voice as being subconscious: she 'noticed' her voice changing and 'noticed' that she had begun to rely on the vocal feminisation techniques more, rather than making a conscious effort to employ the techniques she had learned. In terms of singing, on the other hand, actively working with a vocal coach to androgynise her voice, and to work on her falsetto and mixed voice, indicates a more conscious effort to adapt her voice in this context.

Maria also briefly describes the effects of hormones on her voice. She notes that estrogen will not have as large an effect on her voice as testosterone would for a transmasculine person, but describes some ways in which she believes oestrogen to have affected her voice. Specifically, the feeling of a loss of muscle mass around her throat, which impacts her ability to achieve particular voice qualities at the lower end of her pitch range:

> "That reduction of muscle mass is something that I think I can feel. With regard to my voice as well, with regard to the resonance, [...] realistically, it's not as easy for me anymore to tap into that very bassy resonant voice. And that's happened quite gradually."

Overall, this theme provides an overview of Maria's work to construct her vocal identity, and her attitudes towards the tools available to her for doing this. We see how her aims with regards to her voice change following the surgery, due to an onset of vocal dysphoria. She is seen to consistently position herself against vocal feminisation surgery, and instead opts for learning to control her vocal apparatus in new ways, with help from a vocal coach. In singing, she crafts her sung vocal identity in a very active and conscious way, through setting goals and pursuing these with regular voice training. She also constructs a spoken vocal identity, but in contrast to singing, she typically describes these vocal changes as subconscious.

6.4.4 Theme 4: Voice in Context

Another theme that recurs in the interviews is Maria's differing comfort levels when speaking or singing in the presence of different groups of people, which results in her employing her voice in different ways depending on the context. In terms of singing, she discusses her differing comfort levels in solo compared with group settings, and in singing with exclusively trans people compared with a more general choir. In terms of speech, she comments on the different changes she makes to her voice when speaking with people she has known for a long time, new people, speaking to herself, and speaking when on dates.

Maria talks about employing her singing voice in different ways when singing in solo performance, compared to singing with other people. For example, in the first interview, she describes how she alters the way she projects her voice depending on whether she is singing in a solo or group setting, aiming for a much softer sound when singing low notes in her solo practice, compared with when singing with a choir, where she feels more comfortable allowing herself to "blare like a trombone." She also generally feels comfortable allowing herself to sing at the lower end of her pitch range in a group setting, compared with a solo setting. Singing solo is naturally much more visible and vulnerable than singing with a group, where the other singers provide some cover and it is unclear to the audience exactly which voices come from which people.

This carries over into her differing comfort levels in singing with a choir composed just of trans singers, compared with a choir where this is not the case. In the second interview in particular, Maria talks about her discomfort when singing in a "group of very binary singers" where the audience may have "a certain expectation [that] the girls are singing this, the guys are singing that." This discomfort in singing with "very binary singers" is not purely related to audience perception, however. She describes how a key aim of singing with other trans people is to create a setting where trans singers could feel at ease exploring aspects of their voice that might usually cause them dysphoria. In her own experience, singing with other trans people allows her to feel more comfortable accessing the notes at the lower end of her range, both because her voice is hidden within this 'vocal tapestry', and because she gains comfort from being surrounded by people who share a very similar experience to her.

> "When I'm singing with [a group of trans people] that's always been an intention of ours that, regardless of how you feel about your voice, because you are kind of lost in this trans vocal tapestry, you can play and you can explore where you feel comfortable, and you can access parts of your voice that might otherwise normally cause you dysphoria."

Maria describes how she values her time at a higher rate for situations where she feels less comfortable singing, stating that she now only performs with choirs that are not all-trans when they are highly paid.

When it comes to speech, she also describes using her voice in different ways in different contexts. Specifically, varying her voice quality, in ways which she demonstrates as she speaks to the interviewers, depending on the interlocutor(s):

> "My voice is around here quite naturally. When I'm going on dates and things, maybe it goes a little bit lighter. With friends, especially with friends that I've been around for a long time who've known me from before transition, my voice tends to settle into a much more kind of old school pattern for me, but with everybody else, and to myself, when I'm speaking to myself, it hovers around here. So it's lightened over time."

Again, the way she talks about these changes, with 'my voice' as the subject of active sentences, indicates that the differences in her speaking voice in these different contexts happen subconsciously, rather than her making a conscious effort to sound a particular way when talking to particular people. This is in contrast to her singing voice, where she appears to make more conscious decisions around how to use her voice in different contexts.

Overall, Theme 4 details the differing comfort levels that Maria experiences when using her voice in different contexts. Like the participants in the study of Khoo & Ilbury (2024), she feels much more comfortable in queer-friendly settings, and this impacts the way she chooses to use her voice. Similar also to the participants in the study of Stewart et al. (2020), she expresses feeling less pressure to maintain a stereotypically 'feminine' voice when in the company of other trans people or people with whom she has strong rapport or who she has known a long time. The phenomenon of style shift is rarely explored with trans people, and is a term seldom applied to singing. As I do not have data from Maria in the company of different types of interlocutors, I do not speak further on style shift here, but argue that the present research suggests that further sociolinguistic research with trans (singing) participants could provide interesting contributions to theories around style shift, and invites further research on this topic.

6.5 Discussion

Since an inductive bottom-up approach was taken for the thematic analysis, whereby the themes are driven by the data rather than by a specific research question, the results of this Study serve largely as a descriptive account of the lived experience of one participant before and after undergoing facial feminisation surgery, with a specific focus on her voice. Through this process, a number of impacts of facial feminisation surgery on the voice were identified. Firstly, Maria described temporarily feeling a lack of control over her voice quality when singing in the highest part of her range after the surgery. Secondly, she experienced an unexpected onset of vocal dysphoria following the surgery, despite feeling confident in her relationship with her voice prior to this. Finally, she perceived an incongruence between her face and voice following the surgery, which she was concerned could impact her success as a professional singer. These are all potential impacts of the surgery on the voice that would be helpful for a person considering undergoing facial feminisation surgery to be aware of, but which are not captured by acoustic or perceptual analysis. The level of nuance here would also not have been captured by a standardised questionnaire like the VHI or Transgender Woman Voice questionnaire. In this way, these findings highlight the importance of using qualitative research methods when investigating the impact that surgery can have on the voice.

The extent and pace with which Maria's relationship with her voice changed throughout this study highlights an inherent difficulty in carrying out this type of research. People with vocal dysphoria are unlikely to want to record their voice, and to consent to their voice data being used for research purposes. The potential for facial feminisation surgery (in this case) to cause an onset of vocal dysphoria emphasises the importance of highlighting the concept of informed and continuous consent to participants in similar studies in the future.

As well as discussing the ways in which the surgery affected her relationship with her voice, Maria also explicitly states that she is making a conscious effort to change her singing voice following the surgery. This is crucial information for the correct interpretation of the results from Study 1, as it shows that the observed acoustic changes to her voice following surgery are not purely due to the surgery changing the shape of her vocal tract. Additionally, there was a noticeable change in Maria's mood across the interviews. Affect is known to be reflected in the voice, and so it is also possible that some of the acoustic changes identified in Study 1 are caused by Maria's different emotional states across the recording sessions. It remains difficult to disentangle which specific acoustic results may be caused by articulatory changes made by Maria and which (if any) are caused by the alteration of the vocal tract characteristics. This is left to future research, which could possibly approach this question with more detailed knowledge of the exact surgical procedures undergone by the participant(s). The present Study simply demonstrates that the acoustic impact of physical changes caused by facial feminisation surgery cannot be investigated in isolation.

Chapter 7

General Discussion and Conclusions

This thesis has presented three Studies which approach the question of how the voice is implicated in facial feminisation surgery for one particular participant, Maria, in three different ways. Study 1 presented an acoustic analysis of Maria's speech and singing voice before and after surgery, finding significant changes in her formants, voice quality and /s/ production following surgery. Study 2 explored whether these differences were perceptible to naive listeners, finding the differences to be noticeable but only to some listeners and for some stimuli, indicating that the differences are perceptually very subtle. Study 3 consisted of a thematic analysis of interviews with Maria at several stages during her facial feminisation journey, which provided a descriptive account of how she felt the surgery impacted her voice, as well as her relationship with her voice.

7.1 Returning to the Research Questions

To reiterate, this thesis aimed to provide insight into the following research questions:

RQ1 How can facial feminisation surgery impact the voice acoustically?

RQ2 Are acoustic changes to the voice following facial feminisation surgery perceptually noticeable to external listeners?

RQ3 How can facial feminisation surgery affect a person's own perception of, and relationship with, their voice?

Study 1 addressed RQ1, finding a number of statistically significant acoustic changes to Maria's the voice following facial feminisation surgery. Specifically, in singing, F1 showed a significant increase over each consecutive recording session. F3 was significantly lower 3 months after the surgery for both speech and singing, and had increased again by 7 months following the surgery. H1-related voice quality measures all followed a similar pattern to F3 for speech and singing, with the exception of H1-A1 and H1-A3, which showed a significant drop following surgery, and remained low at 7 months following surgery. Centre of gravity of /s/ also significantly changed across the three recording sessions for both speech and singing. For speech, centre of gravity of /s/ showed a decrease at 3 months following surgery compared with before surgery, and increased again to pre-surgery levels by 7-months post-surgery. In singing, centre of gravity of /s/ increased consistently across the three consecutive recording sessions. Study 3 indicated that, while it is possible that some of these changes were the direct effects of the surgery altering the vocal tract, Maria experienced a large change in her relationship with her voice as a result of her new facial appearance, which also caused her to (sub)consciously alter her speech and singing. Therefore, it seems that the surgery impacted the voice both directly and indirectly, causing a number of acoustic changes.[1]

RQ2 was considered in Study 2, which found that listeners rated samples from two different recording sessions as significantly less similar than two samples from the same recording session. The only exception to this was when listeners were asked to compare singing samples from the 3-months post-surgery and 7-months post-surgery recording sessions, which they rated in a similar way to samples from the same session. This indicates that the changes to Maria's voice following facial feminisation surgery were perceptually noticeable to external listeners. There was a large amount of confusion, indicating that the differences were only marginally noticeable, and that they were more noticeable in some stimuli, and to some listeners than others.

Finally, Study 3 targeted RQ3, finding that the facial feminisation surgery caused a substantial change in Maria's relationship with her voice. She felt that, following the surgery, there was an incongruence between her face and her voice. This, combined with the fact that the surgery reduced the amount of dysphoria she experienced in relation to her face, resulted in a substantial increase in the dysphoria she experienced relating to her voice.

Altogether the results presented in this thesis contribute to answering the overarching question, of how the voice can be impacted by facial feminisation surgery. For Maria, the surgery caused a number of acoustically measurable changes to her voice, which were perceptually noticeable some of the time, to some listeners. She described experiencing less control over her voice quality in the highest part of her pitch range immediately following the surgery, and this subsided over time. She also experienced an increase in the intensity of vocal dysphoria, which led to her making a (mostly) conscious effort to change her singing voice and a (mostly) subconscious effort to alter her speech, and changed the ways in which she felt comfortable using her voice in different contexts.

7.2 Implications for Investigation of Impacts of Facial Surgeries on the Voice

In general, there is a need for reform in the methods used to analyse the impact of facial surgeries on the voice. Studies 1 and 3 together provide the strongest evidence for this. Prior studies on this topic tend to conduct acoustic analyses, and if voice self-perception information is gathered it is usually in the form of a questionnaire which is not always appropriate for the context, or which fails to collect crucial information about self perception of the voice following surgery. Results of these self-perception questionnaires are never linked to the acoustic results in discussion. Information on quality of life following the surgery is often collected in separate studies, and also never linked to the acoustic results. Study 1 of this thesis shows that acoustic analysis alone can lead to incorrectly concluding that observed changes caused by surgery are due to the direct effects of the surgery on the vocal tract.

[1] There is also the possibility that some of these changes were caused by some other factor that was not discussed in the interviews. While it seems likely that the majority of the changes were related to the facial feminisation surgery, this possibility cannot be ruled out, since this is a case study.

Study 3 highlights that, in the case of facial feminisation surgery, the procedure can induce a change in a person's relationship with their voice, which can in turn lead to a change in the way they choose to use their voice. Observed acoustic changes then, can be both due to the changing shape of the vocal tract and due to the person making different use of their vocal apparatus, given the changed social context. The emotional impact that facial feminisation can have on a person, including the ways in which it might affect their relationship with their voice, are likely to differ substantially across people. In turn, the ways in which they adapt their voice (or choose not to) following the surgery is also likely to vary. Taken as a whole, Studies 1 and 3 emphasise the need for social factors to be taken into account when interpreting acoustic changes.

Study 2 concerns a different issue in the evaluation of vocal changes brought on by healthcare interventions more generally. Specifically, that acoustic changes which are statistically significant are not always perceptually meaningful. Acoustic results are also not interpretable to people who have a limited understanding of acoustics, which is the majority of people who would be using this information in their decision to undergo the healthcare option in question. In previous work on the impacts of facial surgeries and of gender-affirming speech therapy on the voice, if perceptual information is sought, it is either done using judgements of speech therapists, or using rating judgements for scales such as 'femininity' and 'pleasantness.' While judgements from speech therapists are undoubtedly helpful, the majority of people that a person encounters are not speech therapists, and so experiments with naive listeners give a better idea as to whether people in the life of the person undergoing healthcare will perceive them differently. Studies on the impact of gender-affirming speech therapy on the voice sometimes use perceptual experiments with naive listeners, but tend to involve rating tasks on parameters which are subjective value judgements. Study 2, presented an alternative method for obtaining perceptual judgements from naive listeners, through simply asking for voice quality similarity ratings. Results showed that listeners perceived a subtle difference between pre-surgery and post-surgery recordings for singing, and between all three recording sessions for speech, which provides a much clearer picture of the changes in the voice following facial feminisation surgery than the acoustic analysis.

The case study in this thesis focused on facial feminisation surgery, but the arguments presented here can be applied to the evaluation of potential vocal changes caused by other healthcare procedures. Returning to rhinoplasty, a procedure which is undertaken by cis as well as trans people, this too can affect a person's relationship with their face. People often undergo rhinoplasty for cosmetic reasons. A successful operation, then, can lead to a notable change in appearance, which could lead to an increase in confidence, and a change in the way the person is perceived by others, for example. This alters the social context and can therefore lead to the person altering how they use their voice, both consciously and subconsciously. If the vocal changes are conscious, then an interview on the person's relationship with their voice would reveal this. If subconscious, then other methods would be needed.

I end by reiterating some crucial limitations of the Studies presented here. I do not argue that the methods used in Studies 1, 2 & 3 of this thesis provide a good model upon which to base future investigations of the impacts of surgery on the voice. On the contrary: the Studies presented here were based on data which was collected for an entirely different purpose, and they therefore possess a number of methodological issues, as highlighted throughout the thesis. Most critically, this thesis presents a case study, and so the findings on the effects of facial feminisation surgery on the voice cannot be generalised. Maria experienced minor physical changes to her voice following the surgery, according to her self-perception and the perception of external listeners, but other people undergo more extreme vocal changes following the same procedures. In particular, chondrolaryngoplasty is known to have the potential to severely limit a person's voice (Strickland et al. 2022, Nuyen et al. 2023). Maria experienced a substantial change to her relationship with her voice following surgery, which is also likely to differ hugely for each person, in ways which are not predictable. This thesis therefore simply provides one example of the ways in which the voice can change following facial feminisation surgery, and uses this to argue for reform in the way the impact of surgical procedures on the voice is investigated.

7.3 Conclusions

To summarise, this thesis has presented an investigation of the acoustic, externally-perceived and self-perceived effects of facial feminisation surgery on the speech and singing of one participant. A number of vocal changes were observed post-surgery, including formant, voice quality and centre of gravity of /s/ measurements. Differences in the participant's voice before, 3 months after and 7 months after surgery were perceptually noticeable to naive listeners, though subtle. Interview data with the participant before and after surgery revealed that the surgery had a profound impact on her relationship with her voice, which led to her making a conscious effort to adapt her articulatory patterns. Taken together, these results were used to argue for the importance of taking into account social factors which can impact the voice when investigating vocal changes caused by facial surgeries. There remains much work to do in documenting the effects of facial feminisation surgery on the voice, since this thesis presents only a case study, and many procedures fall under the umbrella term of 'facial feminisation surgery', each of which will affect different people in different ways. It is hoped that this work will serve to motivate future research on the topic, with more participants, so that people considering undergoing the procedures can have a fuller understanding of the likely vocal changes involved.

Bibliography

Acoustical Society of America (2013, reaffirmed 2024), 'Acoustical terminology', ASA S1.1-2013 (R2024).

Adler, R. K., Hirsch, S. & Pickering, J. (2019), *Voice and communication therapy for the transgender/gender diverse client : A comprehensive clinical guide.*, third edition / edited by richard k. adler, phd, ccc-slp, sandy hirsch, ms, ccc-slp, jack pickering, phd, ccc-slp. edn, Plural Publishing Inc., San Diego, CA.

Aires, M. M., Vasconcelos, D. d. & Moraes, B. T. d. (2021), 'Chondrolaryngoplasty in transgender women: Prospective analysis of voice and aesthetic satisfaction', *International Journal of Transgender Health* **22**(4), 394–402.

Altman, K. (2012), 'Facial feminization surgery: Current state of the art', *International Journal of Oral and Maxillofacial Surgery* **41**(8), 885–894.
URL: *https://www.sciencedirect.com/science/article/pii/S090150271200197X*

Altman, K. (2018), 'Forehead reduction and orbital contouring in facial feminisation surgery for transgender females', *British Journal of Oral and Maxillofacial Surgery* **56**(3), 192–197.

Amino, K., Sugawara, T. & Arai, T. (2006), Speaker similarities in human perception and their spectral properties, *in* 'Proc. of WESPAC', Vol. 9.

Avery, J. D. & Liss, J. M. (1996), 'Acoustic characteristics of less-masculine-sounding male speech', *The Journal of the Acoustical Society of America* **99**(6), 3738–3748.

Azul, D. (2013), How do voices become gendered? A critical examination of everyday and medical constructions of the relationship between voice, sex, and gender identity, *in* 'Challenging popular myths of sex, gender and biology', Springer, pp. 77–88.

Azul, D. (2015), 'Transmasculine people's vocal situations: A critical review of gender-related discourses and empirical data', *International journal of language & communication disorders* **50**(1), 31–47.

Azul, D. (2016), 'Gender-related aspects of transmasculine people's vocal situations: Insights from a qualitative content analysis of interview transcripts', *International journal of language & communication disorders* **51**(6), 672–684.

Bakhshaee, M., Jahanian, M., Khazaeni, K., Sobhani, D., Mashhadi, L. & Rasoulian, B. (2021), 'Potential effects of rhinoplasty on voice in professional voice users', *Aesthetic Plastic Surgery* **45**, 2280–2286.

Barreda, S. (2016), 'Investigating the use of formant frequencies in listener judgments of speaker size', *Journal of Phonetics* **55**, 1–18.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015), 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software* **67**(1), 1–48.

Baumann, O. & Belin, P. (2010), 'Perceptual scaling of voice identity: Common dimensions for different vowels and speakers', *Psychological Research PRPF* **74**(1), 110–120.

Bele, I. V. (2007), 'Dimensionality in voice quality', *Journal of Voice* **21**(3), 257–272.

Bencsik, I. (2022), 'Like real people sing: A linguistic analysis of irish singer-songwriter hozier's singing style'.

Boersma, P. et al. (1993), Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, *in* 'Proceedings of the institute of phonetic sciences', Vol. 17, Amsterdam, pp. 97–110.

Boulet, M. J. & Oddens, B. J. (1996), 'Female voice changes around and after the menopause—an initial investigation', *Maturitas* **23**(1), 15–21.

Bowers, J., Tobey, E. A. & Shaye, R. (1985), 'An acoustic-speech study of patients who received orthognathic surgery', *American journal of orthodontics* **88**(5), 373–379.

Boyd, Z. (2018), 'Cross-linguistic variation of /s/ as an index of non-normative sexual orientation and masculinity in French and German men'.

Braun, V. & Clarke, V. (2006), 'Using thematic analysis in psychology', *Qualitative Research in Psychology* **3**(2), 77–101.
URL: *https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa*

Brown, O. (2002), 'Sensations', *Journal of Singing-the Official Journal of the National Association of Teachers of Singing* **58**(3), 229–232.

Bucholtz, M. & Hall, K. (2005), 'Identity and interaction: A sociocultural linguistic approach', *Discourse studies* **7**(4-5), 585–614.

Buckley, D. P., Dahl, K. L., Cler, G. J. & Stepp, C. E. (2020), 'Transmasculine voice modification: A case study', *Journal of voice* **34**(6), 903–910.

Bultynck, C., Cosyns, M., T'Sjoen, G., Van Borsel, J. & Bonte, K. (2021), 'Thyroplasty Type III to lower the vocal pitch in trans men', *Otolaryngology-head and neck surgery* **164**(1), 157–159.

Bultynck, C., Pas, C., Defreyne, J., Cosyns, M., den Heijer, M. & T'Sjoen, G. (2017), 'Self-perception of voice in transgender persons during cross-sex hormone therapy', *The Laryngoscope* **127**(12), 2796–2804.

Carew, L., Dacakis, G. & Oates, J. (2007), 'The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals', *Journal of voice* **21**(5), 591–603.

Cartei, V., Bond, R. & Reby, D. (2014), 'What makes a voice masculine: Physiological and acoustical correlates of women's ratings of men's vocal masculinity', *Hormones and Behavior* **66**(4), 569–576.
URL: *https://www.sciencedirect.com/science/article/pii/S0018506X14001639*

Cave, R. & Bloch, S. (2021), 'Voice banking for people living with motor neurone disease: Views and expectations', *International Journal of Language & Communication Disorders* **56**(1), 116–129.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/1460-6984.12588*

Cecchin, L. M. (1992), 'Waits v. Frito-Lay, Inc. and Tracy-Locke, Inc. 978 F. 2d 1093 (9th Cir. 1992)', *DePaul-LCA J. Art & Ent. L.* **3**, 88.

Çelik, Ö., Çelik, A., Ateşpare, A., Boyacı, Z., Çelebi, Ş., Gündüz, T., Aksungar, F. B. & Yelken, K. (2013), 'Voice and speech changes in various phases of menstrual cycle', *Journal of Voice* **27**(5), 622–626.

Claros, P., Blebea, C., Pujol, M. C., Claros-Pujol, A. & Claros, A. (2021), 'Voice after rhinoplasty. An important question for the opera singers', *Journal of Voice* **35**(4), 614–617.

Coleman, E., Radix, A., Bouman, W., Brown, G., de Vries, A., Deutsch, M., Ettner, R., Fraser, L., Goodman, M., Hancock, A., Karasic, D., Knudson, G., Leibowitz, S., Monstrey, S., Motmans, J., Nahata, L., Reisner, S., Richards, C., Schechter, L., Tangpricha, V., Van Trotsenburg, M., Winter, S., Ducheny, K., Allen, L., Azul, D., Bagga, H., Başar, K., Bathory, D., Berg, D., Berli, J., Bluebond-Langner, R., Bouman, M.-B., Brassard, P., Byrne, J., Capitán, L., Cargill, C., Carswell, J., Chang, S., Chelvakumar, G., Dalke, K., De Cuypere, G., de Vries, E., Den Heijer, M., Dhejne, C., D'Marco, A., Edmiston, E., Edwards-Leeper, L., Ehrbar, R., Ehrensaft, D., Eisfeld, J., Elaut, E., Erickson-Schroth, L., Feldman, J., Fisher, A., Garcia, M., Gijs, L., Green, S., Hall, B., Hardy, T., Jacobs, L., Janssen, A., Johnson, K., Klink, D., Kreukels, B., Kuper, L., Kvach, E., Malouf, M., Massey, R., Mazur, T., McLachlan, C., Morrison, S., Neira, P., Nygren, U., Oates, J., Pagkalos, G., Patton, J., Phanuphak, N., Rachlin, K., Rider, G., Ristori, J., Robbins-Cherry, S., Roberts, S., Rodriguez-Wallberg, K., Rosenthal, S., Sabir, K., Safer, J., Scheim, A., Seal, L., Sehoole, T., Spencer, K., St. Amand, C., Steensma, T., Tilleman, K., Vala, L., Van Mello, N., Veale, J., Vencill, J., Wesp, L., West, M. & Arcelus, J. (2022), 'Standards of care for the health of transgender and gender diverse people, version 8', *International journal of transgender health* **23**(1), S1–S259.

Coulter, B. (2017), "Singing from the heart": Notions of gendered authenticity in pop music, *in* 'The Routledge Research Companion to Popular Music and Gender', Routledge, pp. 267–280.

Coupland, N. (1980), 'Style-shifting in a Cardiff work-setting', *Language in society* **9**(1), 1–12.

Dacakis, G., Davies, S., Oates, J. M., Douglas, J. M. & Johnston, J. R. (2013), 'Development and preliminary evaluation of the transsexual voice questionnaire for male-to-female transsexuals', *Journal of Voice* **27**(3), 312–320.

Daniel, R. K. (2018), 'The preservation rhinoplasty: A new rhinoplasty revolution'.

Dart, S. N. (1991), *Articulatory and acoustic properties of apical and laminal articulations*, University of California, Los Angeles.

Davies, S. & Goldberg, J. M. (2006), 'Clinical aspects of transgender speech feminization and masculinization', *International Journal of Transgenderism* **9**(3-4), 167–196.

de Araujo, C. M., Schroder, A. G. D., de Araujo, B. M. d. M., Cavalcante-Leão, B. L., Stechman-Neto, J., Zeigelboim, B. S., Santos, R. S. & Guariza-Filho, O. (2019), 'Impact of orthodontic-surgical treatment on quality of life: A meta-analysis', *European Journal of Orthodontics* **42**(3), 281–289.
URL: *https://doi.org/10.1093/ejo/cjz093*

de Leeuw, J. R., Gilbert, R. A. & Luchterhandt, B. (2023), 'jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments', *Journal of Open Source Software* **8**(85), 5351.
URL: *https://doi.org/10.21105/joss.05351*

Delattre, P. (1951), 'The physiological interpretation of sound spectrograms', *Pmla* **66**(5), 864–875.

Deuster, D., Matulat, P., Knief, A., Zitzmann, M., Rosslau, K., Szukaj, M., am Zehnhoff-Dinnesen, A. & Schmidt, C.-M. (2016), 'Voice deepening under testosterone treatment in female-to-male gender dysphoric individuals', *European archives of oto-rhino-laryngology* **273**(4), 959–965.

DiCanio, C. (2013), 'Spectral moments of fricative spectra script in Praat'.
URL: *https://www.acsu.buffalo.edu/ cdicanio/scripts/Time_averaging_for_fricatives_4.0.praat*

Dickinson, K. (2002), 'Believe'? Vocoders, digitalised female identity and camp, *in* 'Pop Music and Easy Listening', Routledge, pp. 343–357.

Duan, Z., Fang, H., Li, B., Sim, K. C. & Wang, Y. (2013), The NUS sung and spoken lyrics corpus: A quantitative comparison of singing and speech, *in* '2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference', pp. 1–9.

Eckert, P. (2012), 'Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation', *Annual review of Anthropology* **41**(1), 87–100.

Eckert, P. & McConnell-Ginet, S. (2003), *Language and Gender*, Cambridge University Press.

Esposito, C. M. (2012), 'An acoustic and electroglottographic study of White Hmong tone and phonation', *Journal of Phonetics* **40**(3), 466–476.

Esposito, C. M. & Khan, S. u. D. (2020), 'The cross-linguistic patterns of phonation types', *Language and linguistics compass* **14**(12).

Eyssel, J., Koehler, A., Dekker, A., Sehner, S. & Nieder, T. O. (2017), 'Needs and concerns of transgender individuals regarding interdisciplinary transgender healthcare: A non-clinical online survey', *PloS one* **12**(8), e0183014–e0183014.

Fairbanks, G. (1960), *Voice and Articulation Drillbook*, Harper & Row.

Fischer-Jørgensen, E. (1967), 'Phonetic analysis of breathy (murmured) vowels in Gujarati', *Annual Report of the Institute of Phonetics University of Copenhagen* **2**, 35–85.

Fitch, W. T. & Giedd, J. (1999), 'Morphology and development of the human vocal tract: A study using magnetic resonance imaging', *The Journal of the Acoustical Society of America* **106**(3), 1511–1522.

Fleischer, M., Pinkert, S., Mattheus, W., Mainka, A. & Mürbe, D. (2015), 'Formant frequencies and bandwidths of the vocal tract transfer function are affected by the mechanical impedance of the vocal tract wall', *Biomechanics and modeling in mechanobiology* **14**, 719–733.

Foroughian, M., Khazaeni, K., Haghi, M. R., Jahangiri, N., Mashhadi, L. & Bakhshaee, M. (2014), 'The potential effects of rhinoplasty on voice', *Plastic and reconstructive surgery* **133**(2), 109e–113e.

Forrest, K., Weismer, G., Milenkovic, P. & Dougall, R. N. (1988), 'Statistical analysis of word-initial voiceless obstruents: Preliminary data', *The Journal of the Acoustical Society of America* **84**(1), 115–123.
URL: *https://doi.org/10.1121/1.396977*

Fox, R. A. & Nissen, S. L. (2005), 'Sex-related acoustic changes in voiceless English fricatives'.

Fry, D. B., Abramson, A. S., Eimas, P. D. & Liberman, A. M. (1962), 'The identification and discrimination of synthetic vowels', *Language and speech* **5**(4), 171–189.

Gelfer, M. P. & Tice, R. M. (2013), 'Perceptual and acoustic outcomes of voice therapy for male-to-female transgender individuals immediately after therapy and 15 months later', *Journal of voice* **27**(3), 335–347.

Gibson, A. M. (2019), 'Sociophonetics of popular music: Insights from corpus analysis and speech perception experiments'.

Graham, F. (2022), 'To T or Not to T: The transmasculine singing voice on hormone replacement therapy', *Voice and Speech Review* **16**(2), 180–199.
URL: *https://doi.org/10.1080/23268263.2022.2038349*

Guarro, G., Brunelli, F., Rasile, B. & Alfano, C. (2019), 'Effects and changes on voice after rhinoplasty: A long-term report', *Plastic Surgery* **27**(3), 230–236.

Gursky, A. K., Chinta, S. R., Wyatt, H. P., Belisario, M. N., Shah, A. R., Kantar, R. S. & Rodriguez, E. D. (2024), 'A comprehensive analysis of genioplasty in facial feminization surgery: A systematic review and institutional cohort study', *Journal of Clinical Medicine* **14**(1), 182.

Hammarberg, B. (2000), 'Voice research and clinical needs', *Folia phoniatrica et logopaedica* **52**(1-3), 93–102.

Hancock, A., Childs, K. D. & Irwig, M. S. (2017), 'Trans male voice in the first year of testosterone therapy: Make no assumptions', *Journal of speech, language, and hearing research* **60**(9), 2472–2482.

Hancock, A. & Helenius, L. (2012), 'Adolescent male-to-female transgender voice and communication therapy', *Journal of communication disorders* **45**(5), 313–324.

Hancock, A., Krissinger, J. & Owen, K. (2011), 'Voice perceptions and quality of life of transgender people', *Journal of Voice* **25**, 553–558.

Henton, C. & Bladon, R. (1985), 'Breathiness in normal female speech: Inefficiency versus desirability', *Language & Communication* **5**(3), 221–227.

Hirano, M., Kurita, S. & Sakaguchi, S. (1989), 'Ageing of the vibratory tissue of human vocal folds', *Acta oto-laryngologica* **107**(5-6), 428–433.

Hodges-Simeon, C. R., Grail, G. P. O., Albert, G., Groll, M. D., Stepp, C. E., Carré, J. M. & Arnocky, S. A. (2021), 'Testosterone therapy masculinizes speech and gender presentation in transgender men', *Scientific reports* **11**(1), 3494–3494.

Holmberg, J., Linander, I., Södersten, M. & Karlsson, F. (2023), 'Exploring motives and perceived barriers for voice modification: The views of transgender and gender-diverse voice clients', *Journal of Speech, Language, and Hearing Research* **66**(7), 2246–2259.

Hughes, C., Brown, G., Ma, N. & Dibben, N. (2024), Acoustic effects of facial feminisation surgery on speech and singing: A case study, *in* 'Processings of Interspeech 2024'.

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S. & Newman, C. W. (1997), 'The voice handicap index (VHI) development and validation', *American journal of speech-language pathology* **6**(3), 66–70.

Jongman, A., Wayland, R. & Wong, S. (2000), 'Acoustic characteristics of English fricatives', *The Journal of the Acoustical Society of America* **108**(3), 1252–1263.

Junior, C. N. V. & de Medeiros, A. M. (2022), 'Voice and gender incongruence: Relationship between vocal self-perception and mental health of trans women', *Journal of Voice* **36**(6), 808–813.

Kelly, V., Hertegård, S., Eriksson, J., Nygren, U. & Södersten, M. (2019), 'Effects of gender-confirming pitch-raising surgery in transgender women a long-term follow-up study of acoustic and patient-reported data', *Journal of Voice* **33**(5), 781–791.
URL: *https://www.sciencedirect.com/science/article/pii/S0892199717305489*

Khoo, J. & Ilbury, C. (2024), 'Navigating 'safe' and 'non-safe' queer spaces: A study of style-shifting in singapore', *Journal of Language and Sexuality* **13**(2), 201–226.

Kletzien, H., Macdonald, C. L., Orne, J., Francis, D. O., Leverson, G., Wendt, E., Sippel, R. S. & Connor, N. P. (2018), 'Comparison between patient-perceived voice changes and quantitative voice measures in the first postoperative year after thyroidectomy: A secondary analysis of a randomized clinical trial', *JAMA Otolaryngology–Head & Neck Surgery* **144**(11), 995–1003.

Konert-Panek, M. (2017), 'Overshooting Americanisation. Accent stylisation in pop singing–acoustic properties of the bath and trap vowels in focus', *Research in Language (RiL)* **15**(4), 371–384.

Köster, O., Schiller, N. O. et al. (1997), 'Different influences of the native language of a listener on speaker recognition', *Forensic Linguistics* **4**, 18–28.

Kreiman, J. & Gerratt, B. R. (1998), 'Validity of rating scale measures of voice quality', *The Journal of the Acoustical Society of America* **104**(3), 1598–1608.

Kreiman, J., Gerratt, B. R. & Precoda, K. (1990), 'Listener experience and perception of voice quality', *Journal of Speech, Language, and Hearing Research* **33**(1), 103–115.

Kreiman, J., Gerratt, B. R., Precoda, K. & Berke, G. S. (1992), 'Individual differences in voice quality perception', *Journal of Speech, Language, and Hearing Research* **35**(3), 512–520.

Kreiman, J., Shue, Y.-L., Chen, G., Iseli, M., Gerratt, B. R., Neubauer, J. & Alwan, A. (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.

Kreiman, J., Vanlancker-Sidtis, D. & Gerratt, B. R. (2003), Defining and measuring voice quality, *in* 'ISCA Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis'.

Lã, F. M. B. & Ardura, D. (2022), 'What voice-related metrics change with menopause? A systematic review and meta-analysis study', *Journal of Voice* **36**(3), 438–e1.

Labov, W. (1966), The Social Stratification of English in New York City, PhD thesis, Columbia University.

Lathrop-Marshall, H., Keyser, M. M. B., Jhingree, S., Giduz, N., Bocklage, C., Couldwell, S., Edwards, H., Glesener, T., Moss, K., Frazier-Bowers, S., Phillips, C., Turvey, T., Blakey, G., White, R., Mielke, J., Zajac, D. & Jacox, L. A. (2021), 'Orthognathic speech pathology: Impacts of Class III malocclusion on speech', *European Journal of Orthodontics* **44**(3), 340–351.
URL: *https://doi.org/10.1093/ejo/cjab067*

Laver, J. (1980), 'The phonetic description of voice quality', *Cambridge Studies in Linguistics London* **31**, 1–186.

Le Page, R. (1986), 'Acts of identity', *English today* **2**(4), 21–24.

Lee, A. S., Whitehill, T. L., Ciocca, V. & Samman, N. (2002), 'Acoustic and perceptual analysis of the sibilant sound /s/ before and after orthognathic surgery', *Journal of oral and maxillofacial surgery* **60**(4), 364–372.

Li, K.-P., Hughes, G. W. & House, A. S. (1969), 'Correlation characteristics and dimensionality of speech spectra', *The Journal of the Acoustical Society of America* **46**(4B), 1019–1025.

Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. (1957), 'The discrimination of speech sounds within and across phoneme boundaries', *Journal of experimental psychology* **54**(5), 358–368.

Linville, S. E. (1998), 'Acoustic correlates of perceived versus actual sexual orientation in men's speech', *Folia phoniatrica et logopaedica* **50**(1), 35–48.

Linville, S. E. (2002), 'Source characteristics of aged voice assessed from long-term average spectra', *Journal of Voice* **16**(4), 472–479.

Liu, S., Babel, M. & Zhu, J. (2024), A comparison of voice similarity through acoustics, human perception and deep neural network (DNN) speaker verification systems, *in* 'Proc. Interspeech 2024', pp. 3674–3678.

Löfqvist, A. (1986), 'The long-time-average spectrum as a tool in voice research', *Journal of phonetics* **14**(3-4), 471–475.

Matsumoto, H., Hiki, S., Sone, T. & Nimura, T. (1973), 'Multidimensional representation of personal quality of vowels and its acoustical correlates', *IEEE Transactions on Audio and Electroacoustics* **21**(5), 428–436.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M. & Sonderegger, M. (2017), Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi, *in* 'Proc. Interspeech 2017', pp. 498–502.

McFee, B., McVicar, M., Faronbi, D., Roman, I., Gover, M., Balke, S., Seyfarth, S., Malek, A. et al. (2023), 'librosa/librosa: 0.10.1'.
URL: *https://doi.org/10.5281/zenodo.8252662*

Melara, R. D. & Marks, L. E. (1990), 'Interaction among auditory dimensions: Timbre, pitch, and loudness', *Perception & psychophysics* **48**(2), 169–178.

Mendoza, E., Valencia, N., Muñoz, J. & Trujillo, H. (1996), 'Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS)', *Journal of voice* **10**(1), 59–66.

Mills, M., Stoneham, G., Greener, H. H. M., Barker, M.-J., Van Horn, E. & Retieff, C. C. W. (2021), *Voice and communication therapy with trans and non-binary people: Sharing the clinical space*, Jessica Kingsley Publishers, London ; Philadelphia.

Monks, S. (2003), 'Adolescent singers and perceptions of vocal identity', *British Journal of Music Education* **20**(3), 243–256.

Moore, A. (2002), 'Authenticity as authentication', *Popular music* **21**(2), 209–223.

Mészáros, K., Csokonai Vitéz, L., Szabolcs, I., Góth, M., Kovács, L., Görömbei, Z. & Hacki, T. (2005), 'Efficacy of Conservative Voice Treatment in Male-to-Female Transsexuals', *Folia Phoniatrica et Logopaedica* **57**(2), 111–118.
URL: *https://doi.org/10.1159/000083572*

Narasimhan, S. & Soumya, M. (2020), 'Spectral and cepstral measures of vocal fatigue in indian heavy metal vocalists', *Journal of Indian Speech Language & Hearing Association* **34**(2), 241–246.

Narasimhan, S. & Vishal, K. (2017), 'Spectral measures of hoarseness in persons with hyperfunctional voice disorder', *Journal of Voice* **31**(1), 57–61.

Nemati, M., Tahmasebi, A., Mohajerani, H. & Tabrizi, R. (2019), 'Does open rhinoplasty alter voice quality?', *Journal of Oral and Maxillofacial Surgery* **77**(1), 179–e1.

Nolan, F., McDougall, K. & Hudson, T. (2011), Some acoustic correlates of perceived (dis) similarity between same-accent voices., *in* 'ICPhS', Vol. 17, pp. 1506–1509.

Nuyen, B., Qian, Z. J., Rakkar, M., Thomas, J. P., Erickson-DiRenzo, E. & Sung, C. K. (2023), 'Diagnosis and management of vocal complications after chondrolaryngoplasty', *The Laryngoscope* **133**(9), 2301–2307.

Ohala, J. J. (1984), 'An ethological perspective on common cross-language utilization of F0 of voice', *Phonetica* **41**(1), 1–16.

Omoniyi, T. & White, G. (2006), *The Sociolinguistics of Identity [electronic resource].*, Advances in Sociolinguistics, Bloomsbury Publishing, London.

O'Bryan, J. (2015), '"We ARE our instrument!": Forming a singer identity', *Research studies in music education* **37**(1), 123–137.

Paliwal, K. K. & Alsteris, L. D. (2003), Usefulness of phase spectrum in human speech perception., *in* 'INTERSPEECH', pp. 2117–2120.

Papineau, B. (2020), '"Hooked on celebri[ɾ]y": Intervocalic /t/ in the speech and song of Nina Nesbitt', *Lifespans and Styles* **6**(2), 22–31.

Perrachione, T. K., Pierrehumbert, J. B. & Wong, P. (2009), 'Differential neural contributions to native-and foreign-language talker identification.', *Journal of Experimental Psychology: Human Perception and Performance* **35**(6), 1950.

Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J., Röder, S., Andrews, P. W., Fink, B., DeBruine, L. M., Jones, B. C. & Feinberg, D. R. (2014), 'Vocal indicators of body size in men and women: A meta-analysis', *Animal Behaviour* **95**, 89–99.

Pobloth, H. & Kleijn, W. B. (1999), On phase perception in speech, *in* '1999 IEEE International conference on acoustics, speech, and signal processing. Proceedings. ICASSP99 (cat. no. 99CH36258)', Vol. 1, IEEE, pp. 29–32.

Podesva, R. J. (2007), 'Phonation type as a stylistic variable: The use of falsetto in constructing a persona', *Journal of sociolinguistics* **11**(4), 478–504.

Quené, H. (2007), 'On the just noticeable difference for tempo in speech', *Journal of Phonetics* **35**(3), 353–362.

Quinn, S. & Hancock, A. B. (2023), Guidance for research with trans and gender-diverse people, *in* 'Seminars in speech and language', Vol. 44, Thieme Medical Publishers, Inc., pp. 119–136.

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org*

Rohrich, R. J. & Ahmad, J. (2011), 'Rhinoplasty', *Plastic and reconstructive surgery* **128**(2), 49e–73e.

Rubin, D. L. (1992), 'Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants', *Research in Higher education* **33**(4), 511–531.

Sambur, M. (1975), 'Selection of acoustic features for speaker identification', *IEEE Transactions on Acoustics, Speech, and Signal Processing* **23**(2), 176–182.

San Segundo, E., Foulkes, P. & Hughes, V. (2016), Holistic perception of voice quality matters more than L1 when judging speaker similarity in short stimuli, *in* 'Proceedings of the 16th Australasian Conference on Speech Science and Technology', pp. 309–312.

Sanchez, A., Ross, A. & Markl, N. (2024), Beyond the binary: Limitations and possibilities of gender-related speech technology research, *in* 'IEEE Spoken Language Technology Workshop 2024', Institute of Electrical and Electronics Engineers.

Schwartz, M. F. & Rine, H. E. (1968), 'Identification of speaker sex from isolated, whispered vowels', *The Journal of the Acoustical Society of America* **44**(6), 1736–1737.

Schwarz, K., Cielo, C. A., Spritzer, P. M., Villas-Boas, A. P., Costa, A. B., Fontanari, A. M. V., Costa Gomes, B., da Silva, D. C., Schneider, M. A. & Lobato, M. I. R. (2023), 'Speech therapy for transgender women: An updated systematic review and meta-analysis', *Systematic Reviews* **12**(1), 128.

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G. & Zäske, R. (2014), 'Speaker perception', *Wiley Interdisciplinary Reviews: Cognitive Science* **5**(1), 15–25.

Silverman, E.-M. & Zimmer, C. H. (1978), 'Effect of the menstrual cycle on voice quality', *Archives of Otolaryngology* **104**(1), 7–10.

Song, T. E. & Jiang, N. (2017), 'Transgender phonosurgery: A systematic review and meta-analysis', *Otolaryngology–Head and Neck Surgery* **156**(5), 803–808.
URL: *https://aao-hnsfjournals.onlinelibrary.wiley.com/doi/abs/10.1177/0194599817697050*

Spiegel, J. H. (2011), 'Facial determinants of female gender and feminizing forehead cranioplasty', *The Laryngoscope* **121**(2), 250–261.

Stamets, R. A. (1993), 'Ain't nothin' like the real thing, baby: The right of publicity and the singing voice', *Fed. Comm. LJ* **46**, 347.

Stevens, K. N. & House, A. S. (1955), 'Development of a quantitative description of vowel articulation', *The Journal of the Acoustical Society of America* **27**(3), 484–493.

Stewart, L., Oates, J. & O'Halloran, P. (2020), '"My voice is my identity": The role of voice for trans women's participation in sport', *Journal of Voice* **34**(1), 78–87.

Story, B. H., Titze, I. R. & Hoffman, E. A. (2001), 'The relationship of vocal tract shape to three voice qualities', *The Journal of the Acoustical Society of America* **109**(4), 1651–1667.

Strand, E. A. (1999), 'Uncovering the role of gender stereotypes in speech perception', *Journal of language and social psychology* **18**(1), 86–100.

Strickland, L., Sussman, K. A. & Madden, L. L. (2022), 'Vocal quality complication following chondrolaryngoplasty: A reported case', *Journal of Voice* .

Sundberg, J. & Sataloff, R. (2005), 'Vocal tract resonance', *Vocal health and pedagogy: Science, assessment, and treatment* pp. 169–187.

Sweet, B. & Parker, E. C. (2019), 'Female vocal identity development: A phenomenology', *Journal of Research in Music Education* **67**(1), 62–82.

Szakay, A. (2006), Rhythm and pitch as markers of ethnicity in New Zealand English, *in* 'Proceedings of the 11th Australasian international conference on speech science & technology', University of Auckland, pp. 421–426.

Tanner, K., Roy, N., Ash, A. & Buder, E. H. (2005), 'Spectral moments of the long-term average spectrum: sensitive indices of voice change after therapy?', *Journal of Voice* **19**(2), 211–222.

Titze, I. R. (1989), 'Physiologic and acoustic differences between male and female voices', *The Journal of the Acoustical Society of America* **85**(4), 1699–1707.

Traunmüller, H. & Eriksson, A. (1995), 'The frequency range of the voice fundamental in the speech of male and female adults'.

Trudgill, P. (1983), Acts of conflicting identity: The sociolinguistics of British pop-song pronunciation, *in* P. Trudgill, ed., 'On dialect: social and geographical perspectives', Oxford, pp. 251–265.

T'Sjoen, G., Moerman, M., Van Borsel, J., Feyen, E., Rubens, R., Monstrey, S., Hoebeke, P., De Sutter, P. & De Cuypere, G. (2006), 'Impact of voice in transsexuals', *International Journal of Transgenderism* **9**(1), 1–7.

Van Borsel, G., De Cuypere, R., Rubens, B. & Destaerke, J. (2000), 'Voice problems in female-to-male transsexuals', *International journal of language & communication disorders* **35**(3), 427–442.

Van Lancker, D., Kreiman, J. & Emmorey, K. (1985), 'Familiar voice recognition: Patterns and parameters part I: Recognition of backward voices', *Journal of phonetics* **13**(1), 19–38.

Vicenik, C. (2009), 'Phonation measurements script'.
URL: *http://phonetics.linguistics.ucla.edu/facilities/acoustic/PraatVoiceSauceImitator.txt*

Vilanova, I. D., Almeida, S. B., de Araujo, V. S., Santos, R. S., Schroder, A. G. D., Zeigelboim, B. S., Correa, C. d. C., Taveira, K. V. M. & de Araujo, C. M. (2023), 'Impact of orthognathic surgery on voice and speech: A systematic review and meta-analysis', *European Journal of Orthodontics* **45**(6), 747–763.

Webb, H., Free, N., Oates, J. & Paddle, P. (2022), 'The use of vocal fold injection augmentation in a transmasculine patient unsatisfied with voice following testosterone therapy and voice training', *Journal of Voice* **36**(4), 588.e1–588.e6.
URL: *https://www.sciencedirect.com/science/article/pii/S0892199720302988*

Whitling, S., Botzum, H. M. & van Mersbergen, M. R. (2023), 'Degree of breathiness in a synthesized voice signal as it differentiates masculine versus feminine voices', *Journal of Voice* .
URL: *https://www.sciencedirect.com/science/article/pii/S0892199723001509*

Wirz, S. & Beck, J. M. (1995), 'Assessment of voice quality: The vocal profiles analysis scheme', *Perceptual approaches to communication disorders* pp. 39–55.

Woloshyn, A. (2009), 'Imogen Heap as musical cyborg: Renegotiations of power, gender, and sound', *Journal on the Art of Record Production* **4**.

Xiao, C. C., Luetzenberg, F. S., Jiang, N. & Liang, J. (2020), 'Does nasal surgery affect voice outcomes? A systematic review with meta-analyses', *Annals of Otology, Rhinology & Laryngology* **129**(12), 1174–1185.

Yuasa, I. P. (2008), *Culture and gender of voice pitch: A sociophonetic comparison of the Japanese and Americans*, University of Toronto Press.

Ziltzer, R. S., Lett, E., Su-Genyk, P., Chambers, T. & Moayer, R. (2023), 'Needs assessment of gender-affirming face, neck, and voice procedures and the role of gender dysphoria', *Otolaryngology-head and neck surgery* **169**(4), 906–916.

Zimman, L. (2017), 'Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/', *Language in Society* **46**(3), 339–370.

Zimman, L. (2018), 'Transgender voices: Insights on identity, embodiment, and the gender of the voice', *Language and Linguistics Compass* **12**(8), e12284. e12284 LNCO-0742.
URL: *https://compass.onlinelibrary.wiley.com/doi/abs/10.1111/lnc3.12284*