

THE UNIVERSITY OF CHICAGO

TO SOUND LIKE ONESELF: TRANSMASCULINE VOCAL CHANGE AND
SATISFACTION ON HRT

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE HUMANITIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF LINGUISTICS

BY
SAM TOBIAS KARI GRAY

CHICAGO, ILLINOIS

AUGUST 2023

Copyright © 2023 by Sam Tobias Kari Gray

All Rights Reserved

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES	vii
ACKNOWLEDGMENTS	viii
ABSTRACT	ix
1 INTRODUCTION	1
1.1 Not hermetically sealed: Trans community and terminology	3
1.2 "Transmasculine" and generalization in the literature	10
1.3 In summary	12
1.4 Outline of the dissertation	14
2 REVIEW OF THE LITERATURE	18
2.1 Pitch and F0	19
2.1.1 Defining gendered pitch ranges	19
2.1.2 Habitual pitch and HRT	22
2.1.3 Pitch range and HRT	23
2.1.4 Social influence on pitch	25
2.1.5 Of pitch, passing, and performativity	27
2.2 Characteristics of vowels	29
2.2.1 Vowel formant frequencies	29
2.2.2 Vowel space and resonance	33
2.2.3 Vowel space and the "transmasculine accent"	36
2.2.4 Gender and vocal tract length	40
2.2.5 Additional gendered features	42
2.3 Trans voices and satisfaction	45
2.3.1 What <i>is</i> satisfaction?	45
2.3.2 What leads to satisfaction?	49
3 METHODOLOGY OF THE STUDY	53
3.1 Structure of the study	54
3.2 Participant recruitment	56
3.3 Participant demographics	57
3.3.1 Demographic highlight: Age	59
3.3.2 Demographic highlight: Race	62
3.3.3 Demographic highlight: Neurodivergence	63
3.4 The identity questionnaire	65
3.4.1 Structure of the questionnaire	65
3.4.2 Results of the identity questionnaire	68
3.5 Final considerations	77

4	PITCH AND HRT	79
4.1	Methodology	79
4.1.1	Methods and expectations: Read passage	81
4.1.2	Methods and expectations: Image description	83
4.1.3	Methods and expectations: Pitch range	85
4.2	Initial-session pitch and pre-HRT voice lowering efforts	87
4.3	Change in habitual pitch over time	91
4.3.1	Cross-participant pitch change	91
4.3.2	Individual patterns in pitch change	94
4.3.3	Change in pitch range, ceiling, and floor over time	97
4.4	Overview of task effects	107
4.4.1	MFD and mean-floor crossover	110
4.4.2	Consideration of image description responses	111
4.5	In Conclusion	115
5	VOWEL FORMANTS, VTL, AND HRT	118
5.1	Change in formants (F1, F2, F3) over time	121
5.1.1	Vowel-specific formant change	122
5.1.2	Participant-specific formant change	123
5.1.3	Task-specific formant change	128
5.2	Change in estimated VTL over time	129
5.3	Discussion of vowel formant and VTL effects	132
6	VOCAL SATISFACTION AND HRT	139
6.1	Satisfaction metrics and expectations	139
6.2	Results: General Satisfaction	143
6.2.1	General satisfaction: Participant variation	146
6.2.2	Mixed responses of Q5 and Q7	150
6.3	Results: Internal Perception	152
6.3.1	Internal Perception: Participant variation	154
6.3.2	Participant variation and what it means to speak in a "way"	163
6.4	Results: External Perception	168
6.4.1	Participant variation in passing and perception	170
6.4.2	Participant experiences with misgendering and assumed queerness	178
6.5	Results: Desire for Future Change	184
6.5.1	Participant variation and desire for future change	185
6.6	Conclusion	194
7	CONCLUSION AND FUTURE WORK	198
7.1	Overview of results	198
7.1.1	Pitch	198
7.1.2	Vowel Characteristics	200
7.1.3	Satisfaction	201
7.2	Methodological considerations	202

7.3	Trajectory of future work	202
7.4	Final comments	206
A	RAINBOW PASSAGE	207
	REFERENCES	208

LIST OF FIGURES

2.1	Approximate mapping of the vowel space to formants F1 and F2.	30
4.1	The colorized <i>Cookie Theft</i> image.	84
4.2	Habitual pitch and floor changes over weeks on HRT.	98
4.3	Change in pitch range over weeks on HRT.	99
5.1	Overall formant changes over weeks on HRT.	121
5.2	Vowel-specific formant changes over weeks on HRT.	124
6.1	"General Satisfaction" question response distribution.	147
6.2	"Internal Perception" question response distribution.	154
6.3	"External Perception" question response distribution.	171
6.4	"Desire for Change" question response distribution.	186

LIST OF TABLES

3.1	Summary of participant demographic information.	58
3.2	Identity questionnaire responses, summarized	68
3.3	Final-session identity questionnaire responses	69
4.1	Vowel tokens analyzed from the <i>Rainbow Passage</i> task	82
4.2	Vowel tokens analyzed from the <i>Cookie Theft</i> task	85
4.3	Starting F0 ceiling, floor, habitual pitch, MRP, and MDP for all participants.	89
4.4	Change over time in MRP and MDP.	93
4.5	Change over time in overall habitual speaking pitch.	94
4.6	Change over time in pitch range and MRD.	100
4.7	Change over time in pitch floor.	104
5.1	Change over time in vowel formants overall.	123
5.2	Change over time in vowel formants per speaker.	125
5.3	Vowel sets showing significant change per formant, per speaker.	127
5.4	(F)irst- and (L)ast-session estimated VTL per participant.	130
5.5	Summary of patterns in vowel formant and VTL change.	132
6.1	Change over time in "General Satisfaction" survey responses.	144
6.2	Notable "General Satisfaction" Q1-9 responses by participant.	149
6.3	Change over time in "Internal Perception" survey responses.	152
6.4	Notable "Internal Perception" Q10-14 responses by participant.	157
6.5	Notable "Internal Perception" Q13 responses by participant.	159
6.6	Notable "Internal Perception" Q12-15 responses by participant.	161
6.7	Change over time in "External Perception" survey responses.	169
6.8	Notable "External Perception" Q22-27 responses by participant.	173
6.9	Notable "External Perception" Q28-32 responses by participant.	179
6.10	Change over time in "Desire for Change" survey responses.	185
6.11	Notable "Future Change" Q33-43 responses by participant.	187

ACKNOWLEDGMENTS

I must first express my gratitude to each of my study participants. Thank you for allowing me insight into an extremely personal journey, and for the profound, enlightening, and sometimes difficult conversations we had along the way. Thank you for the honor of speaking with you throughout this year, for your friendship and enthusiasm, and for being yourselves.

Thank you also to my dissertation chair and committee members for their invaluable support during this project: to Jason Riggle, without whose assistance, advice, and technical expertise this project would never have been completed, and to Sharese King and Diane Brentari, whose sociolinguistic expertise and skills in both interview construction and phonetic analysis have guided me throughout my time at the University of Chicago.

This would not have been possible without my friends and colleagues in Hyde Park, nor those who guided me from early interest in linguistics to the culmination of this project. Thank you to my cohort, Aurora Martinez del Rio, Daniel Lam, Matthew Hewett, Mina Giannoula, and Ömer Eren, and to all those who have made my graduate school experience better by being a part of it; to the members of the Conlang Bulletin Board, who sparked my first love of linguistics; to Emily Gasser and Nathan Sanders, who set me on this path in college and have remained intellectual inspirations.

And of course, thank you to all of my friends and loved ones who have supported me in these years: to Pullman Morris & Sword, the Problem-Busters, and Sasha and Silas Argyris for their help in maintaining work-life balance; to Michael Hill, an eternal source of stability and kindness; to Miriam Attia, whose friendship and support was crucial to the completion of this work; to Mollie Wild, whom I am and will always be infinitely bettered by knowing.

Special thanks to the esteemed Lady Disaster Enchilada Gray, whose constant companionship and tremendous fluffiness have been invaluable emotional support. And most special thanks to my mother, Karen Gray, whose acceptance, love, encouragement, and advice has made all of this possible in too many ways to count.

ABSTRACT

Transmasculine individuals undergoing testosterone hormone replacement therapy (HRT) typically experience a lowering of vocal pitch over the course of the first year, which has been linked to increased satisfaction with their voices. This work addresses two concerns: first, that effects of HRT on vocal characteristics other than pitch are understudied; second, that the degree of individual variation in speakers' vocal goals, identities, and circumstances is such that purely-quantitative metrics do not provide a complete picture of how these acoustic changes contribute to overall vocal satisfaction.

This longitudinal study examines change in habitual pitch, vocal range, vowel formant frequencies, and estimated vocal tract length in 18 transmasculine English speakers over approximately the first year of HRT. It further investigates speaker satisfaction through a combination of surveys and qualitative interviews. Speakers exhibited the expected lowering of habitual pitch and floor; most saw no perceptually salient narrowing of range in semitones. Change in formant frequencies was highly variable, with an overall trend towards lower F1 and F2 alongside persistent and unexpected raising of F3. Satisfaction was not guaranteed for speakers following the first year of HRT. However, the reasons cited behind dissatisfaction were largely unique to the individual, contingent on both self- and external perception of vocal gender as well as the acoustic change seen. Work to improve vocal satisfaction in transmasculine speakers must not only take acoustic properties other than pitch into account, but consider the individual circumstances of those who express dissatisfaction.

CHAPTER 1

INTRODUCTION

This dissertation centers around a year-long longitudinal study that examines both the physical acoustic changes in transgender individuals' voices during the first year of testosterone hormone replacement therapy (HRT) and how these changes relate to the satisfaction reported by these individuals with their voices following that year. It focuses on the experiences of transmasculine speakers (broadly speaking, men or non-binary individuals who were assigned female at birth), with emphasis on how speakers' personal satisfaction varies depending on the social factors of gender identity, presentation, and sexuality, how we discuss and measure satisfaction in the context of trans voices, and the necessity of taking into account individual goals rather than assuming homogeneity across speakers.

The intended audience of this work includes not only fellow phoneticians but members of the trans community for whom its findings may be helpful; as such, efforts have been made to make it accessible to those with limited background in either phonetics or sociolinguistics. The primary goals are **(1)** to examine the many variables that go into the construction of trans identities, and how they combine with the physiological effects of HRT to affect the overall comfort a trans person feels with their voice; **(2)** to quantify these effects, particularly those currently understudied; **(3)** to highlight variation in these effects, emphasizing the importance of individual nuance in discussion of vocal changes on HRT and satisfaction.

The participants of this study are a set of 18 transmasculine English speakers, who share their experiences over the first year of HRT through a series of monthly interviews over Zoom. These interviews combine a phonetic analysis of acoustic features of the voice with in-depth discussion of each participant's experiences. Testosterone HRT is known to cause a drop in pitch (Gorton et al., 2005; Azul, 2015b; Ziegler et al., 2018), with this drop being the primary predictor of participant satisfaction after the first year (Nygren et al., 2009; Deuster et al., 2016; Hancock et al., 2017; Bultynck et al., 2017; Hodges-Simeon et al., 2021).

Along with change in pitch on HRT, participants' vowel formants (F1, F2, and F3) are considered in analysis; existing work on these features in trans speakers has suggested that their production is linked as much to social variation as to biologically-grounded differences in the vocal tract (Zimman, 2012; Papp, 2012). However, no work yet has examined these features in the context of satisfaction. Ultimately, this work aims not only to better understand how social identity and vocal features are linked in trans speakers, but to aid those speakers in achieving greater comfort with their voices. To this end, the study examines possible sources of dissatisfaction, as well as how individual satisfaction is linked to differing goals and identities.

I argue throughout this work against the homogenization of transmasculine experiences in discussion of trans voices. I present evidence that speakers' individual goals regarding their voices vary greatly, and that while drop in mean speaking pitch may be a major predictor of satisfaction, it is far from the only contributor. There is no universal transmasculine desire for one's voice to drop to a pitch within a particular cisgender (i.e. non-transgender) male average range; one man's comfortably androgynous style may be overwhelmingly feminine for another speaker, and vice versa.

Many speakers after a year of HRT still express physical discomfort (Azul, 2015b) or dissatisfaction (Cosyns et al., 2014; Ziegler et al., 2018) with their voices, even after having reached a cis-male-average pitch (Söderpalm et al., 2004; Ziegler et al., 2018), and any discussion of discomfort must take this variation into consideration. This work corroborates these findings; vocal change does not come at the same rate nor to the same extent for all speakers, speakers are not universally entirely satisfied with their voices by the end of the first year, and the reasons for this are as numerous as the speakers themselves.

1.1 Not hermetically sealed: Trans community and terminology

The terminology used throughout this work is based that preferred within the trans communities with which the author has interacted over the course of this study, as well as the personal requests of participants regarding preferred terms and vocabulary. There have historically been, and continue to be, an enormous variety of words and phrases used to describe trans experiences, and we can fully expect the preferred terms to change over time, even as they now vary between individuals and groups. To discuss how factors such as sexuality and presentation relate to vocal characteristics, it is necessary to examine not only the choices of terminology used in this dissertation, but the context behind these choices and how boundaries may be drawn, broken, and re-drawn between different facets of identity.

In particular, we must question the utility of a distinction between identity, presentation, and sexuality, how these ideas overlap, and how to account for their fluidity over the course of transition. We must also examine how they are treated by trans speakers as non-categorical and as non-mutually-exclusive; for example, one participant in this work's longitudinal study, Kel, consistently describes himself as both a *trans man* and a *dyke* (the latter being a reclaimed slur and self-identifier used predominantly by lesbians). I refer frequently in this work to *trans*, *transmasculine*, and *queer* communities and individuals; each of these terms, their history and context, will here be defined and discussed.

The word *transgender* has been used in community spaces, academic contexts, and legal proceedings since the 70s, and was added to the Merriam-Webster dictionary by 1974¹. Initially, it was used predominantly for women who were assigned male at birth and did not desire medical transition. Though now understood commonly as an umbrella term for all whose innate sense of gender identity does not correspond entirely with their sex as assigned at birth, this definition is neither unanimously accepted within trans communities nor

1. *Transfeminine* was added in 1985, and *transmasculine* over 20 years later in 1999, each corresponding to their first attested usages; we thus see terminology reflecting a focus on trans individuals assigned male at birth (over those assigned female) found in both discriminatory legislation and corresponding activism.

historically consistent. The deemphasis of medical transition, however, remains; in modern usage *transgender* does not necessarily imply the presence of physical gender dysphoria² rather being "a broad category typically used to represent any individual whose gender-related identification or external presentation either violates conventional conceptualizations of male or female or mixes different aspects of male and female role and identity" (Diamond and Butterworth, 2008, p365).

The individuals placed within that category have shifted over time; though emerging from within trans communities themselves, *transgender* has at times been ascribed externally by others rather than as a means of self-identification. David Valentine's 2007 work *Imagining Transgender* describes these individuals as *transgender-identified* (i.e. assigned the label *transgender* by external groups due to perceived gender non-conformity, while not necessarily considering themselves as such – for example, gay individuals) and illustrates several examples of how terminology as broad as *transgender* may "simultaneously carry, enable, and restrict meaning" (Valentine, 2007, p19). Fundamental to this work, and at the core of this dissertation, is the recognition that individual experiences of gender are not universal. Shared circumstances between butch lesbians and transgender men, for instance, can lead some individuals to identify with elements of both communities, while other trans men explicitly and firmly distance themselves from lesbian spaces as they relate to no part of womanhood; both possibilities are represented among participants in this study.

The history of *transgender* as a term cannot be discussed without tracing its development as an alternative for the word *transsexual*, which was popularized by Harry Benjamin's *The Transsexual Phenomenon* (1967) and which generally referred to those who desired specific forms of medical transition – particularly surgery. The defining of *transsexual* around this

2. Gender dysphoria here refers to distress experienced due to a disconnect between innate gender identity and physical sex characteristics (e.g. a high-pitched voice may be a source of dysphoria for trans men) as well as with gendered social experiences (such as the use of particular pronouns). It is defined in medical and legal contexts via diagnostic criteria in the DSM-V (American Psychiatric Association, 2022), though some trans individuals reject this as continued pathologization, redefining and reframing dysphoria in various ways. (Johnson, 2019; Beischel et al., 2021)

component of transition led eventually to its use in a predominantly medical, and often pathologized, context.³ For many decades, it was a term that referred only to those who sought these specific procedures, which themselves were available only to those who presented as either masculine, straight (exclusively female-attracted), binary trans men, or feminine, straight (exclusively male-attracted), binary trans women.

These strict criteria ostensibly allowed doctors to distinguish between those who were sincere in their desire to transition and those who would later regret such a decision⁴. Doing so, however, excluded all those unable to meet requirements that notably were not placed upon cisgender patients desiring comparable procedures. This created a double standard, articulated by Zimman who points out that "Despite the fact that a non-trans woman might regret getting breast implants, she is not required to pay for a year of therapy beforehand, for instance, to show that she is certain about her decision. Instead, she is treated as capable of making this decision on her own because she is choosing to enhance her gender normativity rather than diminish it." (Zimman, 2012, p10).

Similarly, these initial standards ignored the existence not only of non-binary individuals, but also of gay trans men and lesbian trans women – or, at least, denied that they could benefit from medical transition. Given the reported experiences of innumerable trans individuals since then, not to mention the final interview session of the longitudinal study I conduct here, in which fifteen out of eighteen participants indicated they were "somewhat" or "extremely" likely to identify themselves as gay (and eight as bi- and/or pansexual), this sentiment can be quite firmly contested.

3. The World Professional Association for Transgender Health (WPATH) is an organization developed from Benjamin's work, providing Standards of Care (SOC) describing conditions to be met before individuals may request medical assistance in transition (Coleman et al., 2022). The SOC requirements, though since substantially revised, have historically prescribed certain behaviors for trans people; these included standards of presentation, such that trans individuals who did not adequately exhibit their gender's expected behavior, attire, vocal style, etc. could be denied care.

4. The most recent research in this regard puts the rate of regret for gender-affirming surgeries at <1% for trans men and women alike (Bustos et al., 2021).

While these standards have since been updated to relax these requirements, there remain certain unofficial expectations towards the behavior and presentation of individuals seeking medical transition. This is a reflection of broader expectations that, even outside of medical contexts, trans people should intend to "pass"; that is, to resemble and be assumed as a cisgender member of their gender. Significant emphasis is placed on passing, both by those trans individuals who desire to do so (either due to personal desire to be viewed as one's actual gender in public, due to discrimination and the need to safely blend into a cisgender-majority population, or both) and academically as a gauge of whether a transition is successful. There has, however, always been a significant population that has no or little desire to pass in this way, even prior to the turn of the century (Kulick, 1999). Trans individuals and researchers alike have noticed "a social drift in the FTM community from straight male-identified stealth men towards openly queer and openly gender-queer or gender-variant individuals" (Papp, 2012, p151), supported by an increase in the past two decades of the acceptance of trans individuals who are gender non-conforming and non-heterosexual (Meier et al., 2013).

Furthermore, it is unclear how non-binary individuals could be expected to "pass" unless this were defined as "being publicly recognized as non-binary", which is only possible in spaces where such identities are recognized. A sentiment found among participants in the current study (see §6.4.1) is the desire to be *able* to pass as male in unfamiliar situations, but also to be distinguishable from a cisgender man in LGBTQ-friendly spaces, particularly vocally. This contrasts with what we may call the "expected trans narrative", a series of experiences associated with a stereotypical "true transsexual" under Benjamin's definition, and which includes the expectation most relevant for this work: that trans people wish to, and should strive to, look and sound indistinguishable from cisgender people of their gender. While crucial to many individuals' comfort, and often to survival, the emphasis placed on passing as the ultimate goal of transition does a disservice to many individuals, and assuming it to be such is an oversimplification of the variety of trans experiences.

As said, the initial conceit of *transgender* was to describe those who lived as the "opposite" gender but who did not pursue medical transition; i.e. one was changing "gender" and not changing "sex". A different, extremely broad definition was given at the turn of the century by Meyerowitz (2002): "...an umbrella term used for those with various forms and degrees of crossgender practices and identifications ... among others, some people who identify as 'butch' or masculine lesbians, as 'fairies,' 'queens,' or feminine gay men, and as heterosexual cross-dressers as well as those who identify as transsexual. The categories are not hermetically sealed, and to a certain extent the boundaries are permeable. The same person might identify as a butch lesbian at one point in life and as an FTM transsexual at another." Meyerowitz (2002, p10).

Others drew different distinctions, for example excluding crossdressers or identifying a strict boundary between sexuality and gender such that a masculine butch lesbian could not also consider themselves a trans man (Califa, 1997). In the last two decades, this definition has continued to develop, with new distinctions drawn and new terms invented. Along with a resurgence in older terms such as *genderfuck* (originating from drag culture in the 1970s (Lonc, 1974)) and *boi* (attributed to rapper Big Boi in 1992, gaining traction in first Black and later trans communities since then), the 2000s have brought us neologisms such as *demiboy* and *cassflux* in a blossoming of linguistic coinage, allowing individuals great specificity in describing their identities; all four of these terms were used by study participants to describe themselves in this work. We see also the use of terms such as *gender non-conforming (GNC)* to describe those who eschew the gender roles ascribed to their sex assigned at birth, but who do not necessarily feel that their gender itself conflicts with that sex.

Based on conversations with study participants, as well as within online and in-person trans community spaces with which I have experience, I observe an "opt-in" component to the modern (2020s) concept of *transgender* as a community term, one that maintains a distinction between sexuality and gender while allowing for nuance on an individual level;

gay men as a group are not considered transgender, and trans women typically draw a strict line between themselves and gay men, but individual gay men may have complex relationships with gender that overlap to some extent with those of transgender women (e.g. preferring *she/her* pronouns). This can extend as far as medical transition, with Diamond and Butterworth (2008) describing many examples of individuals assigned female at birth who have decided to begin HRT but identify their gender as more complex than *man*: some viewed themselves as women, others initially identified as lesbians and continued to do so despite no longer feeling wholly female, and one notably passed day-to-day as male but identified as genderqueer, preferring *she/her* and her (traditionally feminine) birth name.

For some, the term *transgender* carries an inherent expectation that one is transitioning either from male-to-female or female-to-male. This is a common enough interpretation that many who are non-binary, genderqueer, or otherwise outside of this dichotomy do not also consider themselves transgender (Valentine, 2007). This distinction raises important questions in the study of trans voices, which typically have focused on binary speakers; relatively little work has examined non-binary voices in any linguistic capacity, though more recent efforts have been made (Hope and Bradley, 2019; Shefcik and Tsai, 2023). Longitudinal studies examining transition tend to draw a strict line between non-binary individuals and binary trans men, focusing on the latter and making claims about participants' identities, presentations, and vocal goals based on that specific context (Ziegler et al., 2018).

Non-binary as a term does not describe a single homogeneous category, but encapsulates an enormous range of identities, including those who view themselves as between male and female, some combination of them, something else entirely, or entirely devoid of a gender identity. For these speakers, *transgender* may provoke too many assumptions. Words such as *metagender* have been introduced as alternatives, and see use both by individuals and in trans community spaces; this includes one Facebook group from which study participants were recruited, though no participant reported preferring that term.

My own interactions within such spaces further suggest that *transsexual* is typically considered an outdated or pathologizing term akin to *homosexual*, with its use left up to individuals on a personal basis but not adopted as a general term for the community. Some who pursue medical transition such as hormone therapy or surgery do prefer this term, including several participants in this study. *Transgender* has largely taken the stage, with all but two indicating in every session that they were "extremely" or "somewhat likely" to describe themselves as such (§3.4.2); even this, however, is only the second-most consistently preferred term from month to month. The most consistent is a word that sidesteps the question of whether one means *transgender* or *transsexual*, and one which has throughout the decades been used as a shorthand for both: simply *trans*.

For this reason, I use the phrase "trans community" when referring broadly to the real-world and online spaces comprised of individuals who are a different gender than they were assigned at birth, including any who identify with these communities, regardless of desire to pursue transition. I similarly use "trans voices", "trans speech", and so forth, to reference the linguistic qualities of those within this community, with full acknowledgement that it is far from a homogeneous group. This is not unlike *queer*, which perfectly illustrates the melding of gender and sexual identities; gay cis women and bi trans men alike may choose to refer to themselves as *queer*, described by Zimman as "a fluid set of potentially uncategorizable forms of non-normative sexuality" (Zimman, 2012, p1). No distinction is made in the specifics of one's queerness, a freedom from labels that retains a sense of community. This is a sentiment expressed by many of own study participants, some of whom preferred *queer* to the exclusion of all other terms (Chimmy S1|W7)⁵. Again, this is not universal: though *queer* has been largely reclaimed and is now used in academic and political contexts, not all are comfortable with its use for themselves, and due to its associations with non-normative sexuality, some straight binary trans individuals in particular may feel that it does not describe them.

5. Notation here refers to participants' (S)ession number and (W)eeks on HRT.

The only real identifier that I assign to myself is queer, in every sense of the word.

(Chimmy S1|W7)

As such, when referring to the broader community, I will predominantly use the term LGBTQ (Lesbian, Gay, Bisexual, Trans, Queer), except where citations or direct quotes specifically reference topics that use other wording (e.g. "queer studies", "LGBT+ history", and so on). This is done out of the necessity of choosing a single umbrella term of reasonable brevity, and is not intended to exclude groups not explicitly included here (e.g. asexual or pansexual individuals) who may be counted as part of this wider community.

1.2 "Transmasculine" and generalization in the literature

In the title of this work, and in recruiting participants, I explicitly make use of the term *transmasculine* rather than *trans*, referring to individuals who were likely assigned female at birth (AFAB) and who, in the context of this study, were undergoing testosterone (as opposed to estrogen) HRT. *Transmasculine* is a relatively recent term, developed in the late 1990s as an umbrella over both binary trans men and non-binary people who are transitioning from female. It allows for specification of the intended demographic for this study without specifically targeting binary trans men or non-binary people, and is the third-most consistently preferred term by participants.⁶

Existing studies on testosterone HRT's effects are largely based on the assumption that participants are binary trans men; vocal satisfaction is associated with masculinity of pitch, and the desire for a cis-male-passing voice is assumed (Oates and Dacakis, 1997; Descloux et al., 2012; Pettit, 2004). We might call this the "expected transmasculine vocal narrative":

6. Several participants discussed preferring this label overall, as it does not expect them to achieve or specify some target degree of "maleness" or binary identification. It is, however, not without its flaws; in particular, the morphological component "masculine" to many implies a transition neither towards maleness nor away from femaleness, but *towards masculinity*. Such a definition may feel incorrect to, for example, a trans man who prefers highly feminine presentation.

1. The individual realizes he is a man.
2. He desires a cis-male-typical voice.
3. He begins HRT.
4. He develops such a voice without the aid of additional vocal therapy.
5. He goes on to consistently pass as a cisgender man.

This is, however, an oversimplification at every step. First, while we can assume that many of such studies' participants are in fact binary trans men, particularly in works explicitly referencing their self-identification as such, there is no clear dividing line of experiences between binary trans men and non-binary transmasculine people. We could define the difference as one of self-identification, but this ignores experiences of multiplicity and fluidity, as individuals may identify simultaneously with multiple gendered terms, with different ones in different situations, or with none at all. Over the course of the current study, several participants described themselves as both binary and non-binary simultaneously; by its conclusion, a fifth of all participants identified as neither.

Secondly, many transmasculine people either are unable to begin HRT or choose not to do so; such individuals may still seek out vocal therapy.⁷ Of those who do begin HRT, many do *not* develop a cis-male-typical voice, depending on the precise acoustic factors one defines as "cis-male-typical" (which themselves are uncertain, as discussed in §2.1.1), and as many as 16% have been reported as showing "incomplete satisfaction" with their voices (van Borsel et al., 2000; Zraick et al., 2000). Finally, as already discussed, while passing is indeed desired or even critical for many transmasculine individuals, it is far from a universal goal.

There have been attempts to explicitly describe these varied experiences within the transmasculine umbrella. Hansbury (2005) describes three subgroups based on attitudes towards gender, one's body, and society; this categorization is intended only as a framework to contextualize individual nuance, and while I do not make later use of these specific terms, the different experiences they embody can inform how we discuss differences in transition goals

7. Inability may be due to prohibitive cost, lack of access to medical care, contraindications, an un-supportive or unsafe environment, etc.; personal choice is ultimately highly individual, including minimal physical dysphoria or aversion to certain physical changes even as others may be desired.

and satisfaction. Hansbury's "woodworkers" (so named for their tendency to "blend into the woodwork" of cisgender society) are those who "identify themselves as men, minus any transspecific prefixes" (Hansbury, 2005, p246); they are binary trans men who desire medical transition, may not be involved with trans communities, and may prioritize passing as cis. The "transmen", specifically written as a single word, are those who "acknowledge that they live in bodies that persist [...] in being neither female nor wholly male" (Hansbury, p252); they include those who may or may not desire medical transition or whose transitions may not include HRT or surgery, and who may not desire to pass, but for whom maleness or masculinity is an important facet of their identity. Finally, Hansbury's "genderqueers" are those who occupy a "place in the middle, outside any and all boxes. Many eschew gender-specific pronouns, or, conversely, they embrace ambiguity" (Hansbury, p258).

We might summarize these groups very broadly as those who view themselves as "male", "transitioning or having transitioned towards male", and "not female", three different framings of transmasculine experiences; all deserve to have their individual vocal goals and priorities recognized while on HRT. To that end, we must acknowledge several points: **(1)** that there is no single strict dividing line of experiences between binary trans men and nonbinary transmasculine people, or between any of Hansbury's categories; **(2)** that some transmasculine people do not desire a cis-male-typical voice; **(3)** that among those who do, a cis-male-typical voice may not be achieved through HRT alone, resulting in reduced overall vocal satisfaction; **(4)** that among those who do not, vocal satisfaction is *still a concern*, even if not directly related to cis-male-typicality.

1.3 In summary

Various umbrella terms from *transgender* to *trans* to *queer* have arisen to unite different aspects of identity, even as (particularly in the past two decades, with the development of online community spaces) increasingly specific words are being coined to allow for precision

(e.g. Hansbury's categorization, or the aforementioned *cassflux*).⁸ These neologisms appear to reflect a growing individualized sense of identity, critiqued by some as coming at the cost of community cohesion as older members are terminologically othered from younger ones (Green, 2020). Though deserving of more research, the usage and wider social effects of these terms are not the subject of this work; it should be noted, however, that at least among participants of this study these personalized neologisms are employed *alongside* broader terms. A balance may be struck, it seems, between community solidarity and nuance.

We cannot fully divest the characteristics of one community from those of another, nor can we assume that aspects of one subset of the community (e.g. the desire to pass vocally as a cisgender man) will apply to the whole. This applies to sociolinguistic features just as it does to any other characteristic. The ultimate goal of transition is to maximize trans individuals' comfort with their bodies and social realities. This may be framed as reduction of gender dysphoria, or in contrast, the encouragement of gender euphoria (Beischel et al., 2021). One's voice is not only a highly personal component of identity, but a primary means of interacting with others and thus a key feature in gendered perception.

As will be discussed in §2, there is a large body of work on transfeminine voices, with significant resources available for vocal coaches and transfeminine speakers on how to both alter the voice so that is no longer incongruous with gender identity, and to use it so as to better convey that identity in conversation with others. Fewer resources are available for transmasculine speakers, because testosterone HRT typically causes significant physical change in the vocal tract that leads to physiological effects on the voice, corresponding to the changes occurring during puberty for cisgender men; estrogen HRT does not have the opposite effect. There is correspondingly less academic work on transmasculine voices.

8. Even some of these new terms, though specific, may be viewed as an embracing of ambiguity; *cassflux*, used by study participant Wizard, is a neologism describing a fluctuation of indifference towards gender: a very particular kind of vagueness. Participant Ty describes himself as a *demiboy*, conveying a sense of gender that is partially male and partially something else (Green, 2020). Note that these identities are not mutually exclusive with broader or more common terminology; both participants identify as non-binary, with Wizard "extremely likely" and Ty "somewhat likely" to *also* call themselves a man.

Current work on HRT and vocal change has established a drop in habitual speaking pitch, but changes to other acoustic areas, such as pitch range, ceiling, and vowel formants, are still understudied. Each of these elements should further be discussed in tandem with speaker satisfaction; the reported rates of vocal dissatisfaction suggest that even if the first three steps of the "expected transmasculine vocal narrative" are adhered to, the remaining two are not guaranteed. Is it that speakers are not attaining a sufficiently "cis-male-typical" voice, and if so, why? Is this a function of reduced change in pitch, or of other gendered acoustic features? To what extent do speakers have vocal goals other than cis-male-typicality, and how do these goals impact both how their voices change and whether they are ultimately content with how they sound?

No single work can conclusively answer these questions, but to contribute to the discussion, the longitudinal study presented here asks the following: How transmasculine people's voices change acoustically over the first year on HRT, how satisfied speakers are with these changes and in what ways, to what extent these results can truly be generalized across speakers, and what patterns are seen if so.

1.4 Outline of the dissertation

I begin Chapter Two (§2) with an overview of existing literature on gender and the voice, including what acoustic properties are associated with gender in production and perception and how these properties have been discussed in the context of trans speakers. I explore associations of gender with pitch and vowel space, arguing that as these elements are influenced by both physical and behavioral characteristics, an analysis of how they change on HRT must not attribute change wholly to the effects of testosterone, but take behavioral factors into account. This chapter continues to briefly discuss what acoustic properties and voice-related ideas are most salient in trans spaces, including the concept of resonance and the idea of a "trans accent", followed by an overview of how satisfaction has been measured

in trans speakers. I conclude with a broad categorization of the kinds of questions that have been used to investigate satisfaction, critiquing the focus on pitch and passing.

Chapter Three (§3) describes the overall methodology of the longitudinal study, which examines change over approximately the first year of testosterone HRT in a set of 18 speakers, through several vocal tasks and a sociolinguistic interview. It addresses the specific tasks used for analysis of acoustics and vocal satisfaction, means of participant recruitment, and participant demographics. It additionally describes the results of a survey used to track how participant identity informs satisfaction in the voice. I discuss change in identity over time, the prevalence of multiple seemingly-contradictory identities in participants, and how we may take individual identities into consideration when understanding speakers' vocal goals.

In Chapter Four (§4) I begin the analysis of the study's data, presenting the results of three acoustic tasks intended to measure speakers' habitual speaking pitch, floor, and ceiling. I advise caution in certain aspects of study methodology, discussing potential complications from the use of a read-passage task in work with trans speakers; throughout the analysis, we see some speakers, but not all, producing significantly different habitual pitches between a read-passage and image-description task. I further argue against framing pitch change, and particularly range, in terms of hertz, instead supporting a psychoacoustic scale such as semitones to better-capture change in speaker perception rather than production.

The results section of this chapter first discusses speakers' initial pitch measurements and personal vocal goals pre-HRT, continuing on to how pitch changes over time and whether these goals are met. The use of (non-individualized) "target" vocal ranges to measure the success of trans vocal change in existing work is examined and critiqued. Analysis of results incorporates both statistical analyses of change throughout the first year, as well as change within several smaller timeframes, to explore the variety of patterns seen in HRT's effects of pitch. I outline several patterns seen across speakers, focusing in particular on the multiple participants who see effects that are delayed, or who otherwise do not exhibit the com-

mon "cliff-and-plateau" identified in the literature. Overall, while all participants do show lowering of habitual speaking pitch, the rate and extent of this lowering is not consistent.

I then discuss pitch range and the question of vocal range narrowing, the occurrence of which on HRT is currently contentious. Pitch floor drops in parallel with habitual speaking pitch for most, but not all, speakers, while ceiling lowers more drastically, resulting in an apparent narrowing of range in hertz; conversion to semitones, however, reveals little to no significant difference for the majority of participants. Discussion of speakers' interview responses is included to contextualize these changes, in particular those of the few who did see reduction in range, or who expressed difficulty singing over the course of the study.

Chapter Five (§5) similarly presents the results of the analysis of the vowel space, including both the vowel formants F1-F3 and an estimation of speaker vocal tract length. It describes expected patterns of overall formant lowering, if we assume HRT to have a consistent physical effect on the vowel space. The results do not corroborate this; there is instead a general lowering of F1 and F2, which is not consistent either across all vowels or all speakers, and a strong general raising of F3. Estimated vocal tract length also does not appear to be affected by time on HRT, with several speakers instead showing a shortening effect due to the lowered F3. This suggests a strong behavioral influence on the vowel space.

Finally, Chapter Six (§6) discusses the questionnaire used to evaluate satisfaction. It presents results for each question, examining speakers' overall happiness with their voice, efforts made to speak in a particularly-gendered style, self-perception of vocal gender, how their voices are perceived by others, and to what extent they desire further change or vocal training. Statistical analyses show patterns of change over time, including a general increase in satisfaction as well as perceived vocal maleness/masculinity; however, qualitative discussion of interview responses show that these results must be taken in context, as the impact of perceived maleness/masculinity on satisfaction is speaker-specific. I focus in particular on those who showed relatively low satisfaction, and how they articulated their concerns.

The conclusion (§7) summarizes these results in greater detail, describing the overall findings in habitual speaking pitch, ceiling, floor, range, vowel formants, vocal tract length, and the various metrics of satisfaction. It briefly explores how these factors might interact, emphasizing that work to improve satisfaction in trans speakers must first improve awareness of the different patterns of change seen, of what acoustic features may impact how their voice is perceived, and of how their individual identities inform the extent to which HRT affects satisfaction. Some speakers do see different, or reduced, vocal change compared to others; others see substantial change, but still express some level of dissatisfaction which may or may not be improved by vocal training or time. I conclude by looking forward to future work, describing additional analyses to be done of the data collected here as well as further questions prompted by these results.

CHAPTER 2

REVIEW OF THE LITERATURE

In the following chapter, I examine several acoustic properties of the voice which vary by gender, including both in production (where speakers of a different gender tend to have different vocal characteristics) and in perception (where these characteristics influence assumptions made by others about a speaker's gender). I begin with properties of pitch and the vowel space, being the features on which the current study is focused, exploring their relevance to gender and to trans speakers specifically. This includes an overview of methods and results of existing research on change during the first year of HRT. I then briefly touch upon other gender-associated features that are not the focus of this study, such as intonation and use of creaky voice, followed by an overview of work on trans voices and satisfaction, including how satisfaction is defined, how it is measured, and what we have seen so far.

It must be noted that no characteristic of the voice described here is solely linked to gender; there are innumerable other social factors (e.g. region, race, class, age, and so on) which may correlate with each of these features. Some of these elements will be discussed here where relevant, while others will be mentioned in the context of the demographics of this study's participants (§3). I refer repeatedly and separately here to *maleness* and *masculinity* in the voice, and argue that drawing a distinction between the two is useful, such that what is associated with *masculinity* in cisgender men may (instead or in addition) have the effect of indexing *maleness* in transmasculine people. However, this is not without issue; he must take care not to draw an artificial boundary between *maleness* and *masculinity*, or for that matter between gender and sexuality. I can provide no single, unique definition for either *maleness* or *masculinity*¹; if a speaker (or listener) claims something is a "masculine trait" (or vocal characteristic), or if such a trait was identified as such in the relevant literature (e.g. by being common among male speakers), I consider it as such.

1. How this study's participants *individually* define "sounding male/masculine" is discussed in §6.

This is critical to discussion of these features as *indexing* gender, rather than denoting it (see Bucholtz and Hall 2004; Lawson 2020 for further discussion). The tendencies described here as varying between genders do not do so purely due to biological differentiation, but because speakers themselves socially define categories of gender to which speech characteristics are then attributed. For this reason, even when discussing biological or anatomical differences leading to speech differentiation, I refer to *gender* rather than *sex*; as observed by Eckert, "The correlations of sex with linguistic variables are only a reflection of the effects on linguistic behavior of gender – the complex social construction of sex" (Eckert, 1989).

2.1 Pitch and F0

The most noticeable difference between the typical voices of cisgender men and women is pitch, represented acoustically by fundamental frequency (F0), the rate at which vocal folds vibrate during speech. There is evidence that pitch is the single most relevant vocal feature in determining speaker gender (Gelfer and Mikos, 2005); it is thus of particular importance to trans speakers, for whom voice is instrumental to how they are gendered by others.

Difference in pitch between physical sexes is largely biological; higher levels of testosterone, whether introduced during puberty or via HRT during transition, cause lengthening and thickening of the vocal chords that generally results in a drop in pitch (Fitch and Giedd, 1999; Gorton et al., 2005; Evans et al., 2008; Irwig et al., 2017). Cisgender women also undergo a lowering of pitch during puberty, though to a lesser extent than men (Whiteside et al., 2002). However, as with all gendered characteristics of the voice, pitch is not purely biologically determined.

2.1.1 *Defining gendered pitch ranges*

Despite the salience of pitch to gender perception, the range of inter-personal variation means that there there is no universal average range for habitual pitch between genders.

The approximate cross-linguistic cisgender female habitual pitch is between about 180-224Hz (Baken and Orlikoff, 2000; Davies and Goldberg, 2006; Gelfer and Mordaunt, 2018), with significant differences depending on speaker age² Similarly, while average cis male pitches vary from 107-129Hz (ages 21-54, Baken and Orlikoff 2000) to 100–140 Hz (Gelfer and Mordaunt, 2018), work on individual languages shows means from as low as 100–106Hz (from a corpus of British English, Nolan et al. 2006) to as high as 163Hz, 165Hz, or 186Hz (Polish, Punjabi, and Urdu respectively; see French et al. 1998, Demenko et al. 2014); direct comparison between languages has also shown differences, with speakers of German and English producing lower habitual pitches than speakers of Bulgarian and Polish (Demenko et al., 2014).

This variation in age and language, among many other factors, makes it impossible to define universally relevant "target" pitches or ranges for trans speakers who seek to alter how their voices are gendered; a 20-year-old cis-passing Punjabi-speaking trans man may have a much higher pitch than a 50-year-old cis-passing speaker of British English. However, such target values can be convenient abstractions, providing clear (if somewhat arbitrary) goals for such speakers, and are frequently used in discussion of strategies, therapies, and surgeries; this is particularly the case in the context of vocal feminization (Gelfer and Mordaunt, 2018).

Work with transfeminine speakers comprises the vast majority of research on transgender voices. While testosterone hormone replacement therapy results in a thickening of the vocal folds comparable to the effects on cisgender men during puberty, estrogen therapy does not have the reverse effect, and vocal feminization methods are thus significantly more established (van Borsel et al., 2000; Damrose, 2009; Papp, 2012; Zimman, 2012; Cosyns et al., 2014; Azul, 2015b; Davies et al., 2015; Ziegler et al., 2018). The ultimate goals of such strategies are not to fit within some cisgender average habitual pitch, but rather to be *heard as female*

2. Gelfer cites individual examples of 143Hz and 147Hz mean habitual pitches for cis women between 60-69 and 40-50 years of age, as well as reports that women over and under 30 typically speak above and below ~200Hz, respectively (Colton et al., 2011).

by listeners; this typically involves achieving and maintaining a target "gender-acceptable pitch" (Gelfer and Mordaunt, 2018) defined not by production but by *perception*. To this end, collating the results of several perceptual studies, Gelfer suggests a range between approximately 150 and 185Hz within which trans women are "correctly identified as female some or most of the time, based on voice alone" (Gelfer and Mordaunt, 2018, p195).

Such targets and ranges have seen limited use in discussion of transmasculine speech, with Azul et al. (2017) citing "failure to specify normative values" as a specific impediment to comparison of transmasculine speakers' vocal function between studies. One crosslinguistic³ meta-analysis of nineteen studies of transmasculine voices, Ziegler et al. (2018), defined its own pitch targets to enable comparison of "success rates" across studies, examining the percentage of speakers who reached a "cisgender male normative" F0 of ≤ 131 Hz (shared with another meta-analysis, Azul 2015b) or a higher "gender ambiguous frequency" of ≤ 185 Hz. This 131Hz value is based on one measurement of the average pitch of cisgender men, described as the "upper border of the normative range for biological males" (Azul, 2015b, p35)⁴. Notably, this is not necessarily a point below which voices are *perceived* as male. The 185Hz value, in contrast, comes from perceptual work; it is the upper bound of the aforementioned "gender-acceptable pitch" defined by Gelfer, based on F0 ranges for transfeminine speakers.

Ziegler's two target pitches are thus predicated on different contexts: cisgender male production and transfeminine perception. This is largely a consequence of a lack of information on gender perception in pitch for trans men; though such targets are useful in comparing across studies, we must be cautious before generalizing gendered pitch perception in trans-

3. Participants in included studies were speakers of English, German, Dutch, and Swedish.

4. This measurement ultimately originates from work on speech disorders in German speakers (Böhme, 1997), and thus in context references normativity not of gender presentation but of unimpaired speech. Additionally, while corresponding with reported average habitual pitch for German speakers (110-120Hz, Jessen 2009; Demenko et al. 2014), this value's generalization to speakers of other languages must be cautioned, as it is far from a crosslinguistic "upper border".

feminine speakers to transmasculine ones. Given the association of gender with acoustic features other than pitch, there is no guarantee that a "gender ambiguous" 185Hz produced trans women is in fact also gender ambiguous when produced by trans men. Very little work has been done with non-binary speakers, though preliminary work by Hope and Bradley (2019) has suggested such speakers to have habitual pitches intermediary between those of cis men and women, regardless of gender assigned at birth.

Crucially, there is no single target pitch below which one's voice is unambiguously read as male. However, various resources used by trans speakers to masculinize or feminize their voices do present pitch categorically. Discrete gendered pitch ranges are a key feature of numerous "pitch analysis" applications, digital tools that instruct one to speak aloud and track pitch as one does so. Such apps as Voice Pitch Analyzer and Voice Tools aid speakers undergoing voice training by reporting average pitch in Hz as well as a visual representation of how this corresponds to the "(traditional) male, female or androgynous pitch ranges" (Nitz and Seek, 2021). These clearly-delineated, often color-coded pitch ranges provide immediate feedback, though it is unknown from where their specific pitch categorizations originate.⁵

2.1.2 Habitual pitch and HRT

Most trans individuals on testosterone reach a mean speaking pitch in the 107-132 Hz range; this change is well-attested, having been the subject of numerous longitudinal works, and as said above is predominantly attributed to the thickening of the vocal tract comparable to that of typical testosterone-dominant puberty (Papp, 2012; Zimman, 2012; Azul, 2015b; Ziegler et al., 2018). The most dramatic change in pitch is typically within the first year, with most lowering seen within the first six months (Damrose, 2009; Deuster et al., 2016; Nygren et al., 2009; Irwig et al., 2017); as such, most medical and sociolinguistic studies on

5. I did reach out to the creators of Voice Tools to inquire about this, but received no response.

HRT vocal change focus on this time frame. Following this, the change typically levels off.⁶

However, Ziegler et al.'s meta-analysis indicates that while pitch does typically lower over the course of testosterone therapy, the extent of this is not universal. Eight of the studies examined acoustically tracked pitch change over the course of HRT, while the remainder relied on participant self-reporting of vocal change. Across those that did measure F0, all participants exhibited a lowering of F0 at least to a Ziegler's "gender-ambiguous" 185Hz, while 79% reached the 131Hz "cis-male-typical" range.

2.1.3 Pitch range and HRT

Pitch floor, measured in Hz, is known to lower over time on HRT alongside habitual speaking pitch (van Borsel et al. 2000; Constansis 2008; Papp 2012; Zimman 2012, among others cited above). Prior work on habitual pitch versus floor predicted that, upon achieving a cis-male-typical habitual pitch, transmasculine speakers typically speak approximately 3-4 semitones above their new pitch floor (Cooksey, 2000; Papp, 2012). Instead, Papp found such speakers' mean-floor distance (MFD) to universally increase over time on HRT, culminating in mean habitual pitches reaching up to 6-13ST above their floor after one year.

The extent to which pitch ceiling lowers alongside floor, and how this affects speakers' overall vocal range, is inconclusive. Some studies report a significant narrowing of range that persists after the first year on HRT (van Borsel et al., 2000; Damrose, 2009; Hancock et al., 2017). These results, as well as the explanation provided for them, have been quoted verbatim in multiple studies: "The pitch range ... was seriously reduced as a result of the hormone therapy. This reduction appears to be the result of a loss in the high tones which is not fully compensated for by a gain in the lower frequencies" (van Borsel et al., 2000, p434).

This potential range reduction has tremendous implications particularly for transmasculine singers, as well as for trans vocal health in general; the loss of one's ability to sing

6. Individuals have reported either further small "drops" or additional gradual lowering over the course of several more years of HRT (Papp, 2012; Ziegler et al., 2018), though longer studies are relatively rare.

is a major concern, and sentiments around one's singing voice on HRT amount to "use it or lose it". Though difficulties singing may be in part the result of a shifting vocal break, hoarseness or cracking, and the need to redevelop muscle memory for pitch-matching, the expectation of reduction in range is the most salient factor mentioned by participants in the current study (Freddy, S1|W1).

It was one of my hesitations to start T, was knowing that I would lose my current singing voice. And there's no guarantee that I would be able to sing in any way that's pleasing after that, or that I would get a range back. [...] That was one of my main hesitations, knowing that I would be giving that up to start T. (Freddy, S1|W1)

However, other studies (Constansis 2008, Papp 2011) show no such narrowing, and in fact argue that the methods previously used in pitch measurement are sub-optimal. The critique proposed by Papp focuses on the use of hertz as the means of measuring pitch range. Hertz, as the standard unit of frequency, is an accurate metric for speech production; when asking how a speaker's F0 changes over time as a physiological result of HRT, it is applicable. When examining *distance between two frequencies*, it is less so. Human perception of frequency is logarithmic, such that a change from 200-100Hz is perceived as a greater distance (\sim G3-G2 on a chromatic scale, exactly one octave) than one from 500-400Hz (\sim B4-G4, about a third of an octave). As such, a substantially smaller "gain in the lower frequencies" is required to compensate for the "loss in the high tones" than measurement in hertz would suggest.

Papp observed that the overall change in floor-pitch distance in semitones over the 12 months of van Borsel's prior studies was "in the range of one semitone for all three subjects, a musically insignificant change that is well within the margin of error for measuring pitch across multiple measurement sessions" (Papp, 2012, p40); among her own participants, two saw a slight decrease in range, while the remaining four saw a significant *increase*. This is supported by the findings of other studies that focused on semitone measurements: Constansis saw a range increase of 5-9ST in one participant group and 7-14ST in another, though it

must be noted that participants were undergoing singing lessons with the researcher at the time, to which their expanding range may also be attributed.

In contrast, work by Damrose (2009) measured only in Hz, reporting "a profound reduction in mean fundamental frequency within 3–4 months of exposure [...] accompanied by a marked lowering and contraction of pitch range which, after almost 2 years, has remained stable" (Damrose, 2009, p3). The participant of this case study experienced a change in ceiling and floor pitches of $338.49 \rightarrow 201.07\text{Hz}$ and $140.26 \rightarrow 90.75\text{Hz}$, respectively. Upon conversion to ST (methods detailed in §4), we can indeed see their overall range decrease ($15.25 \rightarrow 13.77\text{ST}$), but cannot know whether this change in ST is significant. As such, the effects of HRT on transmasculine speakers' overall vocal ranges remains undetermined.

2.1.4 Social influence on pitch

Pitch is not purely biologically determined, but affected by sociocultural associations both with gender and other components of identity (Bucholtz and Hall, 2004; Lawson, 2020). For example, as described in §2.1.1, reported habitual pitch for particular genders can vary wildly cross-linguistically. This is the case even within single studies, where consistent methodology is guaranteed: Van Bezooijen (1996) found that cis female Japanese speakers spoke at a higher average pitch than cis female Dutch speakers, suggesting broader cultural influences. Yuasa (2008) similarly found a downward trend in F0 for all speakers of American English compared to the results comparable studies in the 1960s, but saw no such trend in Japanese.

Yuasa attributed these trends not to some change in the physiology of female speech but to cultural associations with femininity, proposing that the lower F0 for American female speakers in particular indicated the adoption of more masculine speech patterns. An alternative explanation is that feminine speech norm has shifted; i.e. it is not that American women's speech is becoming "more masculine", but rather the baseline for "feminine" is changing (Zimman, 2012). This is a noteworthy distinction, reflective of a pattern found

not only in production but in perception and in academic work: the explicit association of pitch lowering with masculinization - and by extension, pitch raising with feminization. Indeed, raising the habitual pitch is a primary goal of vocal training for trans women and other transfeminine individuals (Hancock et al., 2014; Adler et al., 2019). The success of such training for many speakers (given a definition of "success" as "achieving a voice that is consistently read as female") is further evidence of the socially-determined nature of pitch. Less frequently, vocal training is used by trans men and other transmasculine people to lower speaking pitch either before, instead of, or in addition to HRT; significantly fewer published resources exist discussing safe and effective methods to do so, although online videos and threads providing advice are not uncommon.

Most individuals have wide pitch ranges and conscious control over where within that range they speak; lower pitch, being associated with men and maleness, can be used to index gender and create a performance of greater "masculinity", as well as associated characteristics such as strength, aggression, etc. (Lawson, 2020). Thus, any efforts to achieve a lower-pitched voice may be regarded as reflecting a desire to attain masculinity, or to become "more male" (reflecting Yuasa's interpretation of lowered pitch in female speakers). Ziegler addresses this assumption, describing the "ubiquity in the literature" of the term *masculinization* and acknowledging that it "underscores a 2-category system of gender identity" (Ziegler et al., 2018, p43), though this work still frames success in terms of achieving target pitches based on cis-male-typical voices. This use of *masculinization* in reference to the effects of testosterone is accurate from a medical perspective; testosterone's effects on the body as a whole may be described as such, as it produces effects also seen in cisgender male puberty. However, this is not necessarily universally reflective of trans vocal situations.

Here we must again caution against overgeneralization; efforts to recognize nuance in the broader transmasculine community must acknowledge individuals who would not phrase their goal as *masculinity* of pitch (e.g. Hansbury's "genderqueers") but rather as *ambiguity*,

androgyny, or in entirely non-gendered terms such as "sounding like oneself". That said, for both binary trans men and non-binary individuals who would prefer to be perceived as male, a voice that is cis-male-typical such that they are able to vocally pass as cisgender is a common goal, the achievement of which is highly correlated with satisfaction for that group (Nygren et al., 2009).

2.1.5 *Of pitch, passing, and performativity*

Though cisgender men with higher-pitched voices do exist (§2.1.1), pitch drop due to testosterone remains a male-typical secondary sex characteristic, and its absence can be a source of gender dysphoria. For such speakers, full awareness of masculinity (vocal or otherwise) as socially-defined, and of the fuzzy boundaries between the constructed categories of "male" and "female" voices, cannot alone alleviate the distress caused by their own voices not reflecting an innate sense of gender identity.

For other speakers, vocal distress may arise not directly from a sense of biological mismatching (e.g. that one "should have" experienced testosterone-based puberty and developed a lower voice, but did not), but indirectly from the responses of others to their voice. The external perception of one's voice as male or female (or as any other gender identity recognized by those with whom one is interacting) will inform interlocutors' responses; as the most salient gendered feature, pitch will contribute to whether one "passes" in public, and can determine the pronouns and forms of address to which others default. I will not dwell in this work on the long history of the concept of "passing", its importance and its pitfalls, how it has been weaponized as proof of trans people "deceiving" others as to their identities, how it has been integral in saving the lives of those in unsafe environments, or how in a modern, increasingly (and ever-so-slowly) accepting world, we seem to be seeing more young trans people for whom passing is not a major concern. For a thorough look at passing, what it is, and why people do it, I point towards Anderson et al. (2020).

Participants in this and other studies varied in how their desire for a change in pitch on HRT derived from physical or social dysphoria, from both, or from non-dysphoria-based understandings of trans comfort such as the "gender euphoria" model (Beischel et al., 2021), as well as how internal ("I wish to sound X way to myself") versus external ("I wish others to read me as X") motivations informed their goals throughout transition. For most, both aspects were highly salient, and even those who described high levels of physical dysphoria also placed emphasis on socially passing (Kevin S1|W2).

I want to pass. I want to sound like a cis man. Um, I don't mind being clocked as trans, especially because I live in a very liberal neighborhood in a very liberal city [...] but at the same time, I would rather pass as cis, because that's the way I feel.

(Kevin S1|W2)

While tremendously important to many trans individuals, as an assumed goal passing has several issues, including that it is often situational and may be largely non-applicable to non-binary individuals who do not anticipate their gender ever being correctly assumed by interlocutors⁷. Notably, several participants expressed a desire to pass as male in public, but to be identifiable as trans in specific environments (Elias S12|W55).

I have a desire to definitely pass more as, like, cis and gender-conforming and stuff, in hetero or strange places. [...] My opposite's, like, very drastically in the other direction. Like, I wanna be visibly trans, but [...] I'm not quite where I'm comfortable with that all the time, more just in queer-centered spaces.

(Elias S12|W55)

The emphasis placed on passing highlights an overall theme of performativity in the voice. Speech is inherently performative, informed by our relationships to society and to our

7. Only 12 of this study's 18 participants consistently indicated "Somewhat/Strongly Agree" to the statement "I want my voice to pass as a cis male's"; see §6. Of these, several desired to do so only in situations where being visibly trans could be met with hostility; others expressed a desire to pass "to themselves" (i.e. recognizing oneself as male in the mirror) but did not care how they were perceived by others; still others were non-binary but desired to pass as male "for convenience" (to avoid confusion or confrontation with those unfamiliar with non-binary identities); etc.

interlocutors; this is not remotely unique to trans speakers (and could be argued is part of the basis for the entire field of sociolinguistics), but it is a theme of which trans communities tend to be uniquely aware. This includes both direct performance of gender identity, and indirect expression of personality traits or characteristics that may themselves feel gendered (e.g. awareness of adopting "a higher, softer, tone when I'm speaking to someone who's in a position of authority" -Oz S1|W15).

2.2 Characteristics of vowels

Aside from pitch, the acoustic properties of the voice most frequently associated with gender are those of vowels; these include variation in specific vowel formants, as well as differences that affect the entire vowel space, such as vocal tract length. Work with transmasculine speakers rarely discusses vowel properties, with F0 being the overwhelming focus (Azul, 2015b); those works that do take other features into account focus predominantly on individual vowel formants, particularly in the context of whether we see change in these formants on testosterone as we do in pitch (Papp, 2012; Zimman, 2012).

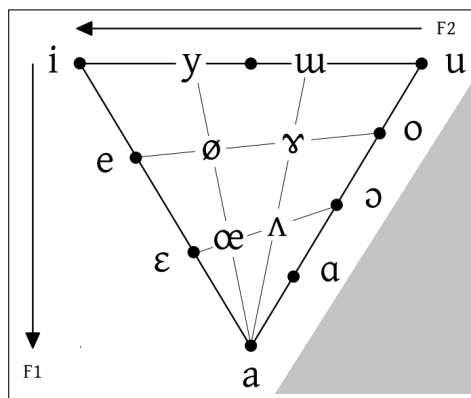
Research on transfeminine voices, and on the practice of voice feminization, dwells more frequently on the vowel space and on how it relates to the concept of "resonance". It is acknowledged both in transfeminine social spaces and by clinical practitioners of vocal therapy that for transfeminine speakers this is equally integral as pitch, if not more so, to gendered perception of the voice (Adler et al., 2019); thus it is from work in this area that we gain our best understanding of *how* vowel quality affects this perception.

2.2.1 *Vowel formant frequencies*

With exceptions for distinctions in length, nasality, and phonation, any given spoken language's vowels are differentiated primarily via formants, which are bands of acoustic energy concentrated around specific frequencies; i.e. "local maximums" or "peaks" of the speech

signal. They are manipulated by changing the position of various speech articulators, most notably the tongue and jaw. Different articulation thus leads to different formant patterns, and therefore to different produced vowels.

Formants are numbered from lowest to highest frequency, with the lowest two (F1 and F2)⁸ the most salient in vowel differentiation. Traditional vowel space mapping describes higher F1 corresponding to lower placement of the tongue in the mouth, and higher F2 to greater frontness (e.g. towards the lips). As such, vowels may be broadly described as *high* (or *close*), *low* (or *open*), *front*, and/or *back* based on the frequencies of these formants. A height-and-backness model is broadly applicable and generally sufficient for understanding vowel production and perception; it is ubiquitous in phonological and sociophonetic research, and most charts of the vowel space are presented in this style (Fig. 2.1).



SOURCE: (Kwamikagami, 2015). CC BY-SA 3.0

Figure 2.1: Approximate mapping of the vowel space to formants F1 and F2.

Though vowel acoustics are well-understood, the precise correlation of tongue placement with formant frequencies has been the subject of some debate. Other models have been proposed; in particular, the laryngeal articulator model (Esling, 2005) instead categorizes vowels as front, raised, and retracted, redefining the importance of the larynx (in combination with oral articulators such as the tongue and jaw) in determining vowel quality and arguing

8. Though labeled in a similar style, the fundamental frequency F0 is not itself a formant.

that "back" vowels are not a discrete category. Furthermore, other vowel-differentiating characteristics have more complicated associations with formant frequency; both lowered F2 and F3 are associated with rounding of the lips, and rhotic or "r-colored" vowels in American English (as in words such as *bird*) exhibit lowered F3 as well.

While the relative applicability of specific models is beyond the context required for this work, it is important to understand the complexity of the vowel space. That said, F1 and F2, being the most salient components of vowel acoustics, are the focus both here and throughout existing sociophonetic literature on vowel space and gender. Female speakers of American English tend to have higher vowel formant frequencies than men (Henton, 1995; Bachorowski and Owren, 1999; Canessa-Pollard et al., 2012; Vorperian et al., 2019). This may be in part the result of physical divergences in the development of the vocal tract, as lower vowel formants are linked to greater length of that tract (see §2.2.4), but socially-motivated differences are nonetheless seen; lower formants are found in boys as young as 4 years of age (Perry et al., 2001), before sexual dimorphism of the vocal tract is apparent.

Perceptually, there is evidence that cisgender female voices are still perceived as such even with an F0 artificially lowered to cis-male-typical range (Gelfer and Bennett, 2013), but that formant frequencies alone are insufficient to change perceived vocal gender when pitch is not altered (Gelfer and Mikos, 2005). On its own, this is strong evidence that factors other than pitch play a role in vocal gender perception; Gelfer & Bennett detail the implications of these results for transfeminine speakers in particular. The artificial manipulation of vowels to have higher formant frequencies, as well as a higher F0, has been shown to increase the rate at which the voice of a cisgender male speaker is perceived as *not-male* (either *female* or *gender-neutral*) to a greater extent than increase F0 alone (Gallena et al., 2018).

Work by Leung et al. (2021) with speakers of Australian English additionally found formant-specific differences in perception; only F2 was shown to be predictive of speaker gender identity (*male* vs. *female*), while F1 and F3 were both instead predictive of speaker

presentation (*masculine* vs. *feminine*)⁹. Though cross-linguistic evidence must be taken with care, Weirich and Simpson (2018) similarly found in German speakers that a lower F1 for the vowel /a/ in particular was associated with greater perception of a voice as *masculine*. Furthermore, while speakers are not typically conscious of vowel space as they are of pitch, cisgender speakers of British English instructed to mimic the speech of the other binary gender consistently produced vowel formants altered as these patterns would predict; female speakers' formants lowered, and those of male speakers raised (Canessa-Pollard et al., 2012).

As with pitch, estrogen HRT has no physiological effect on vowel formant frequency. The effects of testosterone HRT are less certain, with conflicting results from various longitudinal studies over the first year: Zimman (2012) found significant lowering of F1 over time on HRT, but only when summarized over all participants; analysis of each participant's individual formants showed no significant change for any formant, for any participant.¹⁰ Zimman also found no overall change in F2, and an unexpected overall *raising* of F3; he explains these findings as the result of articulatory shift, rather than biological effects of HRT on the vocal tract, and notes that some vowels may shift more than others depending on social context, such as within communities where certain sound changes are taking place.

Papp (2012), in contrast, found a significant lowering of both F1 and F2 in all speakers ($\alpha=0.05$), and an effect on F3 on four of the six, including similar occasional F3 raising. Within speakers, both Zimman and Papp found substantially differing vowel-specific patterns, with Papp reporting overall lowering of nearly every vowel's formants for four out of six speakers, but to varying degrees. Lowering in F2 was consistent across vowels, while

9. Differentiation between "gender identity" and "presentation" here is not absolute. Concepts such as *masculine* and *feminine* may certainly be components of an individual's gender identity, and what constitutes a particularly-gendered "presentation" is entirely culturally informed; I draw this distinction only to highlight nuance in discussion of vocal features. Work with cis speakers rarely concerns itself with vocal *maleness*, but rather with *masculinity*; work with trans speakers tends to conflate the two, which obscures any linguistic stylings we may find associated with, for example, *feminine trans men* or *masculine trans women*.

10. Zimman did use an adjusted α of 0.005 in his analysis, to account for the increased likelihood of false positives as a result of performing multiple comparisons; however, even had he used the standard $\alpha=0.05$, only two of his ten speakers would have showed significant change.

open vowels appeared to show a stronger effect in F1; Papp attributes changes in F1 and F2 partly to changes not only in behavior but in the musculoskeletal structure of the mandible.

Cisgender young men undergoing vocal change during puberty do trend towards a greater overall lowering of F1-F3 than cisgender women over the same age range (Vorperian et al., 2019), and again it is not clear the extent to which this is due to the effects of testosterone directly on the vocal tract via vocal chord thickening (expected in trans speakers) or vocal tract lengthening (possible in trans speakers), indirectly on the vocal tract via changes in height or bone structure (not expected in post-puberty trans speakers), or behavioral adoption of gendered characteristics (expected to varying extents in trans speakers). Overall, whether the changes to F1 and F2 seen in individual speakers is in fact the result of HRT, or the result of non-biologically-motivated changes to speech style, is still unknown.

Aside from individual formants, there are overall patterns in the vowel space as a whole, which may be both predicted by speaker gender and predictive of other's perceptions of gender. The overall size of the vowel space, as measured by distance in F1 and F2 between vowels, tends to be larger in women than men (Henton, 1995), and is correlated with perceived femininity for male speakers (Booz, 2016). This has implications for transgender speakers, as we may then ask whether trans men and other transmasculine people pattern with cisgender men in this regard; i.e. whether femininity in speech (as perceived either by oneself or others) is correlated with larger vowel spaces for such individuals. Depending on whether perceived femininity is desired, this may then contribute to vocal (dis)satisfaction.

2.2.2 Vowel space and resonance

In trans vocal contexts, the vowel space is often framed as one element of the more general idea of "resonance" (Coleman, 1983; McNamara, 2007; Adler et al., 2019). "Resonance" is a term used in literature on voice training, singing, and vocal acoustics, referring broadly to the manner in which the voice passes through and is enhanced by various vocal cavities on

its way out – effectively, the impact of different *resonators*. These resonators amplify the speech signal, affecting vocal quality and timbre in ways that, while somewhat predictable and agreed-upon in vocal literature (e.g. greater chest resonance leading to a "richer" or more "dramatic" sound), do not typically correspond directly to any particular acoustic measurement, nor to specific movements of articulators. As such, they are difficult to quantify.

Descriptions of different "resonance cavities", in particular those familiar as the sources of "chest" versus "head" voice, are found throughout the literature on vocal pedagogy (McNamara, 2007; Adler et al., 2019). The chest, throat, head, and nasal cavity may all be considered resonators, and depending on the precise context, vocal resonance may explicitly refer to the sensation of vibration in these areas while speaking. There is thus immediate physiological feedback for speakers, who can notice when they experience this vibration in different parts of their vocal tract and, potentially, adjust their speech to encourage or discourage it. Being largely informed by such reports of physical sensation, resonance is nebulous to define and measure uniformly in a phonetic context. That said, it serves the very important role of describing non-pitch-related acoustic qualities of the voice in a manner that is physically intuitive to speakers and instructors alike.

Work with transfeminine speakers suggests that gendered associations with these "voices" do exist, and may be related indirectly to variations in pitch and vowel quality resulting from manipulation of these resonators. Discussion, and intentional manipulation, of resonance by any definition involves awareness of the positions of the tongue and jaw (as well as the larynx, and the pharyngeal articulator under Esling's model of vowel articulation), and thus indirectly of the vowel space. Most discussion of resonance in the context of "chest" or "head" voice comes in the context of formal singing training, and the context of voice feminization uses similar language; briefly, resonance in the "head" is associated with vocal femininity, with techniques such as the raising of the larynx applied to increase such resonance, and reduce resonance in the chest (Adler et al., 2019).

The inverse association of chest voice with vocal masculinity is also seen, not only cautioned against for those seeking vocal feminization but sought out by transmasculine speakers. Indeed, techniques borrowing from the language of singing and voice training such as "using chest voice" or "speaking from the chest" is suggested as a method of vocal masculinization (Wizard S1|W1) - and one that appears, anecdotally, to be effective.

I read that one way to sound more masculine or, like, the way that, um, people who grow up [...] as dudes, they'll talk with their chest voice. So I tried to be a little more conscious of that, and I know that's kind of a singer's vocabulary. (Wizard S1|W1)

Among my own participants, some report passing vocally only when they remember to use chest voice; others describe "working on using chest voice" as something that they feel a need to practice, particularly those who after a year of HRT still exhibit some vocal dissatisfaction (§6.5.1). Many, in demonstrating or discussing "masculine" versus "feminine" speech styles, specifically cited chest voice as a major component of the former, and head voice or "speaking in the throat" as a component of the latter. Still others did not associate chest voice directly with masculinity (Arrow S1|W8), but preferred it as a speech style.

I don't really think about masculine and feminine too much. It's usually more like "this sounds nasally" or "speak from your chest". (Arrow S1|W8)

The association of pitch lowering with greater comfort in chest voice corresponds to the description of chest voice as "deeper", with that resonance cavity potentially responding more to lower-frequency pitches (i.e. lower F0, and thus potentially a more cis-male-typical voice). Despite the prevalence of resonance and the language surrounding it in trans spaces, evidence for its significance to gendered speech ultimately remains largely anecdotal. Given the ubiquity of the concepts of "chest" and "head" voice in both pedagogical and casual contexts, their lack of clear definition and ability to be objectively measured has come under scrutiny, with recent work seeking to solve the puzzle by drawing connections between resonance and vowel space.

While little research has been done identifying direct acoustic correlations with resonance in these cavities, work by Cai (2019) has found various locations within the cavities of the head and chest to have different measurable resonance characteristics for different vowels. These broadly correspond to greater high/close vowel resonance in the former, and greater low/open vowel resonance in the latter, such that F1 may be a predictor of resonance location.

Though our understanding of the relationship of resonance spaces to vowel acoustics is still in development, it may ultimately have a practical result in allowing the intentional manipulation of more acoustic properties to take advantage of the intuitive, internally-focused way that resonance is described. Should it be found that some attribute (for example, mean F1) is particularly well-correlated with perceived gender (or, more generally, with overall speaker satisfaction with their voice), such a discovery would benefit from an articulatory framing that uses the language of resonance chambers already so common in vocal pedagogy: it is much easier to help someone to "speak from their chest" than to "lower their F1".

2.2.3 Vowel space and the "transmasculine accent"

Some prior work on the vocal characteristics of transmasculine speakers has identified a vocal style which has been claimed to be distinctive and characteristic of individuals on testosterone HRT; it is typically framed as an undesired vocal quality, either medically as indicative of poor vocal health, or socially as acoustically unpleasant (Constansis, 2008; Zimman, 2012). Constansis discusses several study participants who showed lacked "contour and colour" in the voice, with larynges that did not descend as greatly as those of other speakers. He suggests that this is the result of "immediate high levels of testosterone [which] do not permit the cartilage to adjust gradually. [...] [T]he new male vocal folds can become entrapped within a less-than-adequately enlarged larynx. The resultant voice will sound weak and permanently hoarse and lack the right harmonics." (Constansis, 2008, p16).

The idea of a salient "accent" shared by many transmasculine people is not limited to the

phonetics literature, but expressed within some transmasculine communities and by several speakers in the current study. I first met one participant, Kel, in person through a mutual acquaintance; since he was just beginning medical transition, we came to discuss trans voices, eventually culminating in him asking to join the study. Within the first minutes of that conversation, he had asked verbatim if I as a linguist knew what was up with the "trans guy accent". It is his description of this vocal quality as an "accent" that I have chosen to adopt here, in lieu of complete understanding as to its characteristics. However, though referenced with amusement by Kel, not all descriptions of such an accent are positive; Ty, for instance, expressed frustration towards it, after an incident in which a stranger clocked him as trans at work and claimed they could tell from his voice (see §6.5.1).

I've noticed like, with some trans men at a certain stage, they all kind of sound the same. I don't - I don't really know how to explain it. [...] Transmasc people, they have this sort of tone in their voice, that kind of like – they sound the same. [...] And I think that's what she meant. (Ty S8|W44)

It is Zimman who first links Constansis' "entrapped" voice to the idea of a "trans accent" or transmasculine-specific vocal style. He describes it in more negative terms, reflecting existing stigma around the style, as "a distinctively trans-sounding voice type that community members sometimes talk about as 'tranny voice'" (Zimman, 2012, p193).¹¹ Zimman heavily critiques Constansis' interpretation of this vocal style as the biological product of too much testosterone, too quickly, arguing that these suggestions were made with little evidence and no accompanying statistical analysis of quantitative data. Instead, he posits an explanation based on the vowel space, suggesting this vocal style to be the result of a low F0 accom-

11. The origin of the term "tranny voice" is difficult to determine. The inclusion of the slur *tranny*, which has not been reclaimed except by individuals and which is still used to denigrate transfeminine people in particular, suggests it originated outside the trans community. It is often attributed as an invention of one particular trans social media personality; two of my own study participants specifically cited videos by this individual (here anonymized) as having had a direct influence on community awareness of (and stigma towards) this vocal style. I was unable to locate the exact year that this began, and Zimman's citation of the term's use in trans spaces appears to predate this usage, leaving its true origin unknown.

panying relatively high F1 and F2 formants. If F1 and F2 are not subject to universal biologically-motivated change over time on HRT, but rather change on an individual level depending on social or behavioral factors, this may result in individuals with a more male-typical F0 and more female-typical F1 and F2, producing a characteristic vocal quality not common to either cis men or cis women.

Substantial folk knowledge exists in trans communities surrounding the causes of this phenomenon (also known less pejoratively as the "trans voice" or "trans guy voice"), as well as advice on how it can be avoided. Several years ago, the advice given by community members concerned about developing this vocal style was typically to start HRT at a lower dose, with the reasoning (drawn ultimately from Constansis' speculations) that "people who take higher doses, end up having a lot of voice cracks and possibly having the 'tranny voice' because their body cant get used to the change, as it happens so fast" (sspencerj, 2020).

More recently, however, community sentiment around this voice appears to be shifting. Questions of "How do I avoid the 'trans voice'?" still arise across social media, but community members instead are more likely to advice on the use of chest versus head voice, with greater awareness of voice training techniques (ExistingCactus, 2022). Responses to such questions now tend to agree that it is not universal, not biological (i.e. not the result of too much testosterone, too quickly), and not exclusive to trans men at all. Those participants in the current study who reference this vocal style describe having heard both claims.

"Trans voice" exists because there are multiple components to how a voice sounds, and testosterone only changes one of them. [...] Voice training will get you in the habit of speaking from your chest voice. It usually doesn't need to be as extensive as the voice training trans women do, but it probably will take some purposeful effort.

(ExistingCactus, 2022)

Based on reflections from participants, as well as my own experience in trans community spaces, I believe there has been recent pushback against the stigmatization of trans voices

in general and this style in particular. There is recognition that it may be a natural speech style for many speakers, trans man and non-binary individuals alike (to both of whom the idea of having "failed" to perform vocally by not sounding "acceptably" masculine may be insulting); furthermore, it has been anecdotally reported among cisgender gay men (and thus is not a uniquely trans feature that may potentially out speakers, reducing safety-related concerns). General consensus appears to be that it is a vocal style more akin to an accent, which one may work to change, if desired. Emphasis is placed not on a biological inevitability of HRT, but on individuals' empowerment to train and take control of their voices. In short: "Dude, unlearn that term. There's no such thing as a 'tranny voice.'" (mrsweezydc, 2022).

This greater awareness of the language used by vocal coaches and singers may correspond with greater accessibility to transmasculine-specific vocal training. A more in-depth analysis of how this vocal style has been discussed in social media is required, and may be the focus of future work. With what we know currently, one question still remains: What, acoustically speaking, *is* the "trans accent"? More recent discussion in social media attributes it largely to usage of "head voice", but as said above, this is an extremely nebulous concept. Zimman's speculation regarding a low F0 and relatively high F1 and F2 would conform to this, given aforementioned associations of high vowel formants with femininity or, more broadly, non-cis-maleness of the voice. It is salient enough in trans spaces that it can reportedly be recognized and agreed upon by different individuals, though no study to date has investigated this.¹²

Ultimately, however, any explanation for the "trans voice" remains largely speculation, I must also point out that it cannot be conclusively confirmed that this "entrapped" style is in fact *the same thing* as the "trans accent", though it shares subjective descriptive characteristics. Though we do not investigate either of these phenomena directly in the current study, we can examine the pattern of higher overall vowel formants (e.g. a more "female-typical"

12. Though this remains entirely subjective, my own intuitions do suggest a correlation with sensations of chest resonance. As a vocal style, it can be intentionally produced in a manner recognizable to other trans individuals; I have done so successfully with trans acquaintances, and this does produce a noticeable shift in the sense of physical vibration from chest to throat.

vowel space) alongside lower F0; we may ask whether it corresponds to overall speaker dissatisfaction (as a stigmatized vocal style might suggest), to perception of one's voice as less *male* or *masculine*, or to greater likelihood that one is recognized by others as trans. We must not discount the reality of Constansis' subset of speakers who did experience negative effects on their voice quality or range. Nor, however, can we assume that the "entrapped" vocal quality or "trans accent" is in fact an undesired outcome for all speakers. Though not the primary focus of the current study, the idea of a transmasculine accent deserves consideration, whether stigmatized or embraced.

2.2.4 *Gender and vocal tract length*

In addition to gender differences in vowels on the formant level directly, there is also evidence for physiological gender differences in the vocal tract that may be reflected indirectly in vowel formants. The characteristics most relevant to discussions of HRT and the voice are vocal fold thickness and vocal tract length (VTL). While the former is predominantly related to pitch, VTL is the greatest source of variation in formant frequency after vowel type (Turner et al., 2009). It may be measured accurately via magnetic resonance imaging (MRI), but access to the required equipment for sociolinguistic research is rare.

Instead, estimations of VTL using vowel formants are typically used; higher formants such as F3 and F4 appear less sensitive to articulatory differences and thus are good indicators of VTL, to which they are inversely related (Turner et al., 2009; Lammert and Narayanan, 2015). Both MRI studies and formant-based estimations of VTL found no significant anatomic differences in the male and female vocal tract pre-puberty, but that descent of the larynx and lengthening of the pharynx during puberty causes cisgender men to develop vocal tracts approximately 60% longer than those of cisgender women (Fitch and Giedd, 1999; Turner et al., 2009). Though VTL is predominantly correlated with height (and thus the relatively greater height of cis men compared to cis women is a factor), this

additional descent and lengthening of the male vocal tract appears independent of height, and associated with the onset of testosterone-based puberty.

Manipulation of functional VTL is fundamental to discussion of vocal therapy for trans-feminine speakers, with emphasis placed on a conscious raising of the larynx to reduce this length and thus achieve a higher natural pitch (Gelfer and Mordaunt, 2018). This is a component of the focus on resonance described above. However, while efforts to "speak from the chest" may have the effect of lowering the larynx and thus lengthening the vocal tract for some transmasculine speakers, intentional lengthening of the vocal tract reaches physiological limits without medical or hormonal intervention.

Another method of altering the physiology of the vocal tract is through surgery. There are several types of voice feminization surgery, undergone by many transfeminine individuals for whom vocal therapy does not produce the desired results. Because of testosterone HRT's effects on the voice, voice masculinization surgery is comparatively uncommon; however, in recent years, several techniques have been developed and implemented for individuals for whom testosterone HRT would have no or minimal effect.¹³ It has also been explored as an option for those who have been on HRT but did not experience a satisfactory change in the voice; case studies on such individuals has shown a significant lowering of habitual speaking pitch alongside improved satisfaction. (Saito et al., 2018; Webb et al., 2022)

Though vocal cord thickness and vocal tract length both play a role in determining formant frequencies, work on HRT's effects have focused primarily on the former. While few studies have yet applied formant estimation methods to trans speakers, work by Hodges-Simeon et al. (2021) centered vocal tract length, as estimated by the third and fourth formants (F3-F4), in discussion of satisfaction. They found that the VTL of trans men on average appears intermediate between that of cisgender men and women, with 23% of their trans male participants' estimated VTLs falling outside the range of their cisgender male

13. For example, an individual with androgen insensitivity syndrome (AIS), as in T'Sjoen et al. (2011).

sample. Thus, while significant change in VTL was seen over time on HRT, overall it "may not be sufficient for achieving formant frequencies that are indistinguishable from cis-men" (Hodges-Simeon et al., 2021, p6). Several case studies have also shown changes in estimated VTL over time on HRT, with speakers from the work of Cler et al. (2020) and Buckley et al. (2022) showing an increase from 16.3-16.9cm and 17.0-17.3cm, respectively.

Even this physiological distinction in length may be subject to behavioral manipulation; not only can vocal feminization therapy enable transfeminine speakers to manipulate their larynx and thus achieve more cis-female-typical VTLs, but similar effects may be possible for transmasculine speakers as well, despite the aforementioned physiological limits; in the Buckley case study, the speaker in question saw a further increase in estimated VTL to 19.4cm after "active" laryngeal reposturing, a procedure of muscle training, massage, and stretching. Laryngeal reposturing is a technique used in the reduction of excessive laryngeal tension, which is one of several potential issues reducing vocal quality and comfort in transmasculine speakers (Azul et al., 2021). Dahl et al. (2022) finds that laryngeal massage and reposturing, when applied to transmasculine speakers who had already been on HRT for several years, led to an overall decrease in habitual pitch and increase in estimated VTL, as well as reported subjective vocal masculinity.

2.2.5 Additional gendered features

There are a wide variety of other features of the voice associated with gender, either in production, perception, or both. While none of these are specifically examined in the current study, they must be acknowledged as contributors to how trans voices are perceived.

The first of these are properties of the fricatives /s/ and /ʃ/; the duration, center of gravity, skew, and kurtosis of these sounds have all been shown to differ between genders (Munson et al., 2006; Podesva and Van Hofwegen, 2015; Zimman, 2017). Current findings do suggest transmasculine speakers to have higher fricative centers of gravity than cisgender

male speakers (Papp, 2012; Zimman, 2012, 2017; Pearce, 2019). This is complicated by the association of these properties of /s/ with both the speech produced by some gay men (Munson et al., 2006) and speech that is *perceived* as gay (Mack and Munson, 2012; Kachel et al., 2018); overlap between gay and transmasculine speech patterns is a topic beyond the current work, but a key point of future discussion. The phenomenon of transmasculine individuals passing specifically *as gay men* is one described often among my participants, and perception as such may have a tremendous impact on satisfaction depending on whether the speaker is comfortable with that assumption (which may or may not be accurate).

There is also the matter of creaky voice, also known as vocal fry. In many languages, creak and other phonation types are phonemic (Esposito and Khan, 2020); in others, they are linked to various social characteristics, including gender. Initially described in English as a feature of working-class male speech (Henton and Bladon, 1988), creak has since grown associated with the speech of young American woman (Yuasa, 2010; Podesva, 2014). The physical effects of HRT on the larynx may affect creak, and it has been speculated that creak can contribute to perception of a speaker's voice as male or masculine by proxy, via lowering perceived F0 (Zimman, 2017). However, Pearce (2019) argues against a link between creak and gender, finding no correlation in a comparison of creak usage across cis, trans, and non-binary speakers. Becker et al. (2022) also found no correlation with a similarly diverse sample, though measurement of overall rates saw trans men not on testosterone HRT exhibiting the most creak (39%), with trans men on HRT exhibiting the least (24%). Our understanding of phonation type in transmasculine voices is thus still in its infancy.

Finally, there is the case of intonation and dynamism in the voice. There is a pervasive stereotype of male speech being "monotone", less dynamic and variable in pitch than that of women. This difference is notable because it appears in perception, such that trans women who incorporate more vocal dynamics may be more likely to pass vocally as women (Wolfe et al., 1990), but not necessarily in production; Henton (1995) found no gender differentiation

in vocal dynamism. Other characteristics of intonation are also worth examination, such as certain inflectional patterns associated with women; prominent among these is the use of high rising terminal (colloquially "uptalk", characterized by a sentence-final rising pitch). Some recent work (Hope and Bradley, 2019) has examined intonation specifically in non-binary speakers, finding minor intonation trends that appeared more common in such speakers.

Despite the questionable extent to which intonation is in fact gendered in production, it is one of the most salient aspects of the voice to many trans speakers, including a majority of participants in the current study. Speakers describe using a monotone to attempt to masculinize their voices, and some remarked that they felt they were required to do so in order to be vocally read as male (Southern Boy S1|W3).

I've had people tell me in the past that if I wanted to sound like a man, I had to sacrifice the inf- like, the expressiveness in my voice. And like I mentioned I'm a performer, that's my thing, that's my personality. I'm not a monotone who [*in a monotone*] speaks in this level all the time, [*return to normal voice*] I'm dramatic! I'm really dramatic! And so my hope is that taking testosterone will naturally lower my voice to a level where people still read me as a man without me necessarily having to sacrifice my personality, for lack of a better word. (Southern Boy S1|W3)

Intonation, dynamism, and the gendered associations of various inflectional patterns deserve significant future attention. There are many other features of the voice which have been argued, often more contentiously, to have gendered components, including voice onset time, vowel duration, and other phonations such as breathy voice (Whiteside and Marshall, 2001; Zimman, 2012). Though each of these is outside the focus of the current study, we must again acknowledge that any analysis of speaker satisfaction will necessarily be incomplete without taking all elements of the voice into consideration as a whole.

2.3 Trans voices and satisfaction

Multiple studies have shown that HRT improves general satisfaction and quality of life in transmasculine individuals, and in particular, that it increases satisfaction with one's voice (Nygren et al., 2009; Deuster et al., 2016; Hancock et al., 2017; Bultynck et al., 2017; Hodges-Simeon et al., 2021). In light of this, I pose two questions: How is satisfaction defined, and are there specific acoustic changes on HRT that correlate with it?

2.3.1 *What is satisfaction?*

We consistently see a substantial subset of transmasculine individuals expressing some level of dissatisfaction with their voices following a year on HRT, between 12-17% across numerous studies; we further see up to 25% finding they are still perceived as female over the phone, and up to 31% showing interest in vocal therapy to further masculinize speech (van Borsel et al., 2000; Ziegler et al., 2018; Hodges-Simeon et al., 2021). However, satisfaction has not been consistently defined in the same way across studies; thus, care should be taken when discussing these studies as a group to ensure we are looking across comparable metrics. We must also take care that our definitions of satisfaction do not enforce a particular normative narrative for transmasculine speakers that, for example, ignores the existence of non-binary individuals (see Konnelly 2021 for an overview of nuance and normativity in trans contexts).

Satisfaction has been quantified in several ways. Azul (2015a,b) incorporated interviews, comparing speaker identity to self-perceived vocal gender, perception of gender by others over the phone, and speakers' desired vocal gender attribution. Nygren et al. (2009), Deuster et al. (2016) and Hodges-Simeon et al. (2021) used scales, asking participants to rate from *strongly agree* to *strongly disagree* how satisfied they were with their voice, or whether they agreed with statements to that effect. Van Borsel et al. (2000) incorporated *yes/no* questions (e.g. "Are you satisfied with your present voice?") as well as other rating scale methods (e.g. *never-sometimes-always* for "Are you still addressed as a woman on the telephone?").

Each of these methods has involved gauging agreement with certain phrases or questions. Though some have devised their own metrics, many are based upon either the Transsexual Voice Questionnaire (TVQ) or earlier Transgender Self-Evaluation Questionnaire (TSEQ), initially developed for trans women (Davies, 2006; Dacakis et al., 2013) and later adapted for trans men as the TVQ^{FTM} (Bultynck et al., 2017, 2020; Sirin et al., 2021); this is a collection of such questions that incorporates both self-perception of vocal gender and general vocal (dis)comfort, with particular emphasis on others' perceptions of one's voice and day-to-day impact (e.g. "My voice causes me to lose income"). The TVQ^{FTM} and TSEQ are optimized for work with binary trans speakers, and so more recently the VENI (Voice-related Experiences of Nonbinary Individuals) questionnaire has been proposed as an alternative for that group (Shefcik and Tsai, 2023). To better-discuss how satisfaction is quantified and understood, I categorize questions gauging satisfaction very broadly into several themes: general satisfaction, internal perception, external perception, and desire for further change.

General satisfaction encompasses overall happiness with one's voice, and any reasons for this not directly related to *what gender* one's voice is perceived as. While satisfaction following the first year of HRT is often framed as the extent to which cis-male-typicality of the voice has been achieved, a speaker who does not identify as male and who has minimal physical vocal dysphoria (thus, for whom such typicality may not be a goal) is still capable of dissatisfaction. Examples include sentiments expressed among Van Borsel's participants; by the end of the first year "14 [*out of 16*] subjects were pleased ... Two were pleased but would still have wanted a 'heavier' voice. One subject was not satisfied with his voice because it strained him too much to speak in a low voice." (van Borsel et al., 2000, p430).

This discomfort may be indirectly like to gender (e.g. a "heavier" voice may be interpreted as a desire for greater vocal masculinity) but is not necessarily so. Discussing satisfaction in this way allows speakers to express distress towards their voice that is separate from a gender disconnect: "I sound like a man, *but...*". For example, a common sentiment in

online and in-person trans spaces I have frequented is for transmasculine individuals to be assumed much younger than their actual age, and this extends to the voice; as such, we may see dissatisfaction as a result of sounding young rather than necessarily sounding "female". The work of Bultynck et al. (2017), making use of the TVQ^{FTM}, examined only changes in vocal satisfaction over time, not acoustic features, finding that self-perception of the voice improved over time and most drastically during the first 3 months of HRT, corresponding to the greatest change in pitch (§2.1.2). Finally, work such as that of Azul et al. (2017) finds numerous speakers who felt that their speech was strained or restricted, providing physiological evidence of possible restriction to glottal function as well as reports increased vocal fatigue. Questions relating to this fatigue, sense of cracking, hoarseness, etc. also fall into this category, and comprise much of the TSEQ/TVQ.

Internal perception describes how speakers personally gender their voice. Much existing work has focused on this area, with Hodges-Simeon et al. (2021) asking specifically about participants' happiness with their level of "vocal masculinity". Van Borsel et al. (2000) similarly asked whether participants believe their voices to be "more female than male", "no longer different from a male voice", and other questions with similar wording. These have the benefit of directly targeting speakers' gendered associations with their voice, thus measuring the extent to which they subjectively feel their voice "masculinized" as expected on HRT.

We may also ask how important a specific gendered perception of their voice is to speakers. There is a distinction between "My voice sounds masculine to me" and "I want my voice to sound masculine"; if the latter is answered negatively, we may expect answers to the former to interact differently with satisfaction. Few studies draw a distinction between *female* and *feminine*, or *male* and *masculine*, or explicitly acknowledge vocal androgyny¹⁴. The design of the VENI (Shefcik and Tsai, 2023) is a notable exception, and this work found non-binary speakers to vary in how they desired their voices to be gendered; though not distinguishing

14. That is, few studies *on trans voices*. Perception of masculinity and femininity in the voice has been studied in cis speakers, often alongside sexuality (Kachel et al., 2018).

between between transmasculine and transfeminine individuals, they found 5/10 speakers to desire a voice that was more androgynous, 4/10 more masculine, and 1/10 more feminine.

The TVQ addresses identity linearly asking how speakers gender their voice on a range from "very female" to "very male", with "gender neutral" in the center. This framing is straightforward, but unable to represent a voice that is *male* but *feminine*, or one whose speaker perceives it as both *male* and *female* without necessarily also being *gender neutral* or *androgynous*; even allowing for a middle ground, it is still based upon a binary conception of vocal gender. This will be discussed further in my own analysis of satisfaction (§6), wherein numerous speakers express "mixed-gender" vocal perception that is both *male* and *female*, *binary* and *nonbinary*, or neither. In brief, speaker perception of vocal gender appears not to situate *male* and *female* as opposites, nor *androgynous* as a midpoint between them.

Similarly, some speakers may frame their vocal goals as reflecting an internal sense of self not explicitly tied to gender; one may, for example, desire a pitch that is lower but not necessarily more *masculine*. They may instead have a lower baseline for *feminine* or *androgynous* speech (reflective of Zimman's versus Yuasa's conclusions regarding F0 in modern American women, see §2.1.4). Alternatively, one may view vocal gender as informed wholly by identity, rather than a characteristic of the voice and how it is perceived; e.g. "I am a man, *therefore* my voice is a male one." For such speakers, we might expect internal perception of vocal gender not to be associated with acoustic properties of the voice at all.

External perception, in contrast, relates to how speakers believe their voice is gendered by others. This includes whether speakers think others judge their voice to be *male*, *female*, *masculine*, or *feminine*, as well as whether they believe their voice outs them as trans. This may be distinguished from how speakers' voices are *actually* gendered by others, which requires a perceptual study to accurately determine; such studies have been performed, with work such as that of Owen and Hancock (2011) and Quinn et al. (2021) finding correlations

between self- and listener-ratings of vocal gender in trans speakers.¹⁵ Aside from asking how one's voice is perceived by others on the phone, the TSEQ/TVQ does not draw a distinction between internal and external perception; I suggest that a comprehensive examination of satisfaction in speakers should separate the two, due to cases in which participants read their voices as their desired gender, but are not read as such on the phone or in person: "...three of 16 subjects of the sample survey reported that it still happened that they were addressed as a woman on the street ... in spite of the fact that they themselves considered their voice no longer different from a male voice" (van Borsel et al., 2000, p438).

Desire for further change, finally, is the overall wish for participants' voices to continue changing. This is distinguished from general satisfaction due to the possibility for individuals to be content with their voice but still look forward to further lowering of pitch, a sentiment described Van Borsel et al.'s two subjects quoted above as well as several of my own. It further includes participants' desires to seek vocal therapy.

2.3.2 *What leads to satisfaction?*

The physiological effects of HRT on the voice are limited, and some vocal effects over the first year of HRT may be not the direct result of biological change, but rather due to changes in social situation. Individuals in their first year of transition may also be newly coming out to family or friends, or otherwise altering their social environment. Those who had adopted intentionally feminine speaking styles, to allay suspicion while closeted, may cease doing so.

Additionally, HRT has numerous physiological effects unrelated to the voice, which themselves may improve comfort within one's body, affecting overall satisfaction and thus vocal satisfaction by proxy. As such, it is difficult to determine whether any of those changes lead directly to improved satisfaction, or whether such improvement is the result of greater *overall* comfort with oneself, rather than with one's voice specifically.

15. No work I am aware of has made a three-way comparison of vocal gender in self-perception, listener perception, and *speaker belief of listener perception*; the third is what "external perception" refers to here.

Predictors of satisfaction

In the last decade, these questions have been asked several times. Nygren et al. (2009) found speaker satisfaction to correlate with overall lower F0 values. Deuster et al. (2016) similarly found that the change in fundamental frequency between the start and end of the first year on HRT was the only variable to correlate with increased satisfaction. In the latter case, however, it was not final F0 after 12 months of HRT, but the *magnitude* of F0 change relative to speakers' initial baseline that was the significantly more reliable predictor of satisfaction. This suggests that it is the extent of change, rather than specifically achieving or approaching some cis-male-typical baseline, that predicts final satisfaction with one's voice.

In each of these works, pitch has been the centerpoint; little consideration has been given to acoustic factors other than pitch or to those for whom pitch changes were insignificant. As discussed in §2.1, F0 is the most salient gender-linked component of speech, and because of the consistent effects of testosterone on F0, discussions of transmasculine voices form a five-point narrative as introduced in §1: the individual recognizes a trans male identity, desires a cis-male-typical voice, goes on HRT, sees a drop in pitch, and reaches a cis-male-typical vocal range with no need for further training or intervention and no issue passing vocally as male (Oates and Dacakis, 1997; Pettit, 2004; Descloux et al., 2012). As a result, trans men have been claimed to face little possibility of vocal problems relative to trans women; the sentiment is that "there appears to be considerable agreement among researchers [...] that voice therapy for the female-to-male transsexual is unnecessary because lowering of the fundamental frequency occurs automatically as a result of androgens" (Pettit, 2004, p224).

This narrative is reflective neither of the diversity of trans speakers' vocal goals (with not all speakers desiring either such a drastic drop in pitch, or to pass as cis men, and some who do desire such a drop not exhibiting it) or the reality of their experiences (with many speakers showing dissatisfaction even following pitch drop). Existing work on satisfaction has seen criticism for assuming this narrative; Azul (2015a) took a different approach, examining not

acoustic but social predictors of satisfaction. This work found that improved satisfaction emerged as an interaction between participants' identity, self-perception of vocal gender, and external perception of the same. One participant was dissatisfied with an androgynous perception of their voice because they wished to be perceived as male; another was dissatisfied with a voice that was assumed male, as they sometimes preferred to present as female.

Rates of dissatisfaction

There does appear to be a growing awareness that vocal dissatisfaction following the first year of HRT is not uncommon, that pitch may not be the sole different component to satisfaction, and that satisfaction metrics should acknowledge participants' differing vocal goals. First, we see evidence that as many as 10% of speakers in any given study expressed difficulties maintaining a desired pitch (Ziegler et al., 2018), or were otherwise dissatisfied with their results; though the metrics used to measure satisfaction varied, it was typically described as correlating with participants failing to maintain a cis-male-typical pitch.

Secondly, we see increased acknowledgement over the previous decade that a focus *solely* on pitch is reductive. Azul (2015b) found similar rates of dissatisfaction to Ziegler, aggregating and analyzing longitudinal trials and case studies with greater focus on overall vocal health and comfort. In this work, Azul challenges various assumptions regarding the goals of transmasculine individuals, specifically calling attention to the frequency with which studies on trans voices, pitch change, and satisfaction in particular are inconclusive or contradictory. He particularly notes that the prevalence of issues such as limited pitch lowering, insufficient perceived masculinization, and general dissatisfaction was under-reported. "Only the problem area 'lack of change to voice pitch' as a result of testosterone treatment received substantial attention. All other problem areas were found to be under-researched with reported results to be inconclusive." (Azul, 2015b, p41). The result of this lack of attention is that satisfaction due to any factor other than insufficient drop in pitch is rarely discussed,

and empirical investigations of the root causes of vocal dissatisfaction are sparse. An exception is Hodges-Simeon et al. (2021), who discovered that the VTLs of trans men did increase but often did not reach cis-male-typical norms after the first year of HRT (§2.2.4); based on the salience of the vowel space to perception of gender, they propose that this is a potential source of dissatisfaction.

I strongly agree with the caution suggested by Azul in generalizing both voice change and satisfaction across transmasculine speakers as a uniform group. Language use is at once a performance and a reflection of the self (see Zimman 2015, 2018 for sociological insight into trans voices and language use). We may cite statistics claiming that as many as 12-17% of speakers express dissatisfaction, but ultimately, satisfaction must be largely understood at an individual level. This is not to say that there is no value in describing patterns of (dis)satisfaction across speakers, only that any efforts then made to *improve* satisfaction must be highly individually tailored, taking personal identity and vocal goals into account.

CHAPTER 3

METHODOLOGY OF THE STUDY

This study is in part an attempt to address several issues brought up by authors such as Azul (2015b), regarding both the downplaying of low satisfaction in the literature and the near-exclusive focus on pitch, by furthering this discussion of speaker satisfaction. To do so, it combines acoustic analyses of the changing voice on HRT with numerous quantitative metrics of satisfaction, alongside qualitative sociolinguistic interviews with speakers.

In structure, it is inspired by longitudinal studies which have come before it, intending in part to replicate work done by Papp (2012) and Zimman (2012) while incorporating additional tasks. The study consists of a series of 12 interviews and recording sessions of 18 transmasculine participants during and beyond their first year of HRT. These sessions took place through the Zoom video conferencing software and involved several tasks designed to elicit different data on vocal acoustics and satisfaction. The former included a read passage task, image description task, and pitch range exercise; the latter consisted of two surveys (the "identity" and "satisfaction" surveys) conducted via the Qualtrics platform.

The specific methodology behind each acoustic task will be described in §4 and §5, prior to analysis of each task's results; the satisfaction survey will similarly be discussed in §6. In the following chapter, I provide an overview of the study participants, including means of recruitment and participant demographics. I also briefly explore the responses to the study's "identity" questionnaire and corresponding parts of the interview, which allowed participants to express preferred terminology for their own identities and forms of expression, to be considered in the context of vocal satisfaction.

20 participants were initially recruited for the study; while 18 remained on HRT for the full duration, two were unable to do so. While this work will not include discussion of their vocal situations, I hope in later work to present a more in-depth qualitative look at all participants' interviews, including those of this pair.

3.1 Structure of the study

The 12 recording sessions were spaced slightly over one month apart, with an average of 5.5 weeks between sessions spanning participants' first 12-16 months on HRT. Each session was approximately one hour in length. Participants prior to first session were given instructions on how to use the open source software Audacity to record their own audio, though as described below, not all were able to do so. The first session also included a demographic survey to gather general information on participants' pronouns, general region of residence, known languages, etc. (Table 3.1); the final session concluded with a debrief, in which participants were welcomed to ask questions about the study, its hypotheses and methods.

Prior to each session, participants were sent a link to the paired identity and satisfaction questionnaires on the Qualtrics platform; most completed them in the first few minutes of each session, allowing them to comment on the questions as they answered or to elaborate on their responses, though some preferred to complete the questionnaires outside of the sessions to maximize time that could be spent in-session. Open discussion of the questionnaires then followed. Often no commentary was made or needed, but on occasion considering their responses would prompt participants to mention particular relevant anecdotes or experiences. Participants were then asked generally about their voices, changes noticed since the prior session, and overall vocal health and comfort. Sessions continued with a sociolinguistic interview (Labov, 1981); this is an interview style designed to create conversational flow, rather than a question-and-answer response pattern. It consists of nested modules with which an interviewer may prompt speakers to elaborate further on a topic, but is otherwise largely unstructured, allowing the interviewee to guide the course of discussion.

Later interviews were largely guided by participants' own commentary on their lives and vocal change; the first session was more structured, aiming to provide an overview of speakers' goals for their voices. Though not all topics were discussed in depth by all participants, some example questions are listed below; these are non-comprehensive, as further questions were

prompted by participant's individual responses. This initial discussion was typically followed by the three acoustic tasks, though each task was often separated by further discussion.

- What effects do you hope testosterone therapy will have on your voice?
- How important do you feel like your voice is as a marker of gender identity?
- Have you heard or followed any advice about methods of masculinizing your voice?
- Are there aspects of your identity that you feel impact the way you speak?
- Do you consider yourself a part of a (or "the") trans community?
- To what extent do you interact with other trans people?
- Do you talk differently with different groups of people?
- Do you feel pressure to speak a certain way?

This study took place during the height of COVID-19 pandemic. As such, elicitation could not feasibly be done in-person in a small sound-controlled lab with a shared microphone, as is typical for phonetics work of this nature. Instead, all interviews and data collection were done via the Zoom video conferencing software. How comparable linguistics research performed in this way is to that done in more controlled environments remains a somewhat open question. The concern is primarily one of audio quality, and of how the change of the speech signal through a phone or computer microphone, and possibly then further through Zoom, may affect the measurement of features such as vowel formants.

Current evidence suggests that recordings made locally on Zoom (though not in the cloud) capture acoustic data comparably to those made in-person, particularly in F0 (Zhang et al., 2016). Audio format and compression both play a role here; most physical recorders used for linguistics research store recordings in the lossless *.wav* format, while Zoom's own recordings compress audio to *.mp3* files. Regarding the hardware used, F0 in particular appears equally well represented across various devices (Grillo et al., 2016; Jannetts et al., 2019; Ge et al., 2021). The vowel space overall appear to be adequately captured for acoustic analysis on smartphones, laptops, and more dedicated solid-state recorders (De Decker and Nycz, 2011; Grillo et al., 2016), but cross-device variation increases in higher formants, with greater differences found in F3 and to some extent F2. There is also variation between devices in jitter, shimmer, and vocal harmonics (Ge et al., 2021).

Each speaker was necessarily recording on their own device, and so we may expect this to have introduced some inter-participant variation in measurements of vowel formants, though the extent of that variation in this study cannot be directly quantified. While the interviews were done over Zoom, the audio used in acoustic analysis was typically recorded directly by participants through Audacity; speakers were directed in installing and using the software and saving the audio in the *.wav* format. I, as interviewer, also recorded all sessions using Audacity, though this interviewer-side audio had been passed through Zoom. Not all participants were able to record speaker-side audio, with data from Gareth and Southern Boy in particular relying only on interviewer-side audio and several instances of speaker-side audio being corrupted or lost due to hardware failure. Where possible, however, speaker-side recordings allowed for clearer audio that was not compressed through Zoom.

Participants were instructed to record in quiet spaces, with minimal background noise, preferably on a computer with an external microphone and headset. This was not always possible for all participants, typically due to lack of the necessary hardware; though inter-speaker consistency could not be guaranteed, the recording environment and method was kept as consistent as possible for each individual participant between sessions, and efforts were made to reduce noise and disruption before and during each session.

3.2 Participant recruitment

Participants were collected primarily through social media, with a virtual flyer of the study created and spread initially through Facebook and Discord groups. These groups were trans community spaces, several specifically dedicated to discussing trans voices or vocal training. From here, snowball sampling and sharing of the flyer led to some participants discovering the study through Twitter or other social media. All were required to be at least 18 years old, native speakers of American or Canadian English, and either about to start testosterone HRT or within approximately three months of having started.

3.3 Participant demographics

Demographic information is provided in Table 3.1. Participants' HRT doses and how they changed over the course of the study are also given where disclosed, though evidence suggests that dosage of HRT is significantly less useful than blood levels of testosterone as a predictor of vocal change or other physical effects, as the same dose may be differently metabolized between individuals (Gorton et al., 2005; Zimman, 2012).

Due to the manner in which participants were recruited, and the conduction of the study entirely through Zoom, participant locations regional language variety varied substantially. This range was partly out of necessity; due to COVID restrictions many in-person trans spaces in the Chicago area were closed or not frequented at the start of the study. With in-person recruitment unavailable, I could have tailored online recruitment to look specifically in the Chicago area, or otherwise restricted data collection to a specific region. Instead, although regional variation is a confounding factor in any study of the voice, any common effects across participants despite geographic distance between them would suggest that those effects are not wholly conditioned by the region in which the study happened to take place. This is particularly relevant in discussion of vowel spaces, which vary considerably between regional English varieties but which also show gender-linked properties (§2.2.1).

A total of 20 participants were found, though only 18 remained on HRT for the majority of the study. It is important to note that the experiences of each person could form a case study of much greater length, and that this overview is wholly insufficient to fully capture them as individuals. Over these sessions, I have gotten to know each participant, and now call several friends; among those months of discussion were many conversations about gender, sexuality, and the difficulties and joys of existing while trans. Many have shaped, in part, the direction of my research, or provided insight into new questions to ask other participants. The wisdom, struggles, and laughter shared cannot be fully captured here; however, where relevant, direct quotes from participants will be used to provide context and nuance.

Alias	Pronouns	Age	Location	Ethnicity	Lang.	(S1) HRT Weeks Dose
Arrow	he/they/ she	32	California	Black	Spanish	8 Gel: alternating between 20.25mg/d and 40.5mg/d
ATM	he/him	27	Washington	White	<i>n/a</i>	2 SubQ: 1mL bw (1-4), .3mL/w (5-8), .4mL/w (9-12)
Chimmy	he/him	39	Virginia	White	<i>n/a</i>	7 Gel: 20.25mg/d (1), 40.5 (2); SubQ: .25ml/w (5-7), .4 (8-9), .5 (10-12)
Elias	he/they	25	Texas	White	<i>n/a</i>	0 IM: .2mL/w (2-4), .3 (5-12)
Freddy	he/him	31	Texas	White	French	1 IM: .3mL/w
Gareth	he/they	53	Vermont	White	French, Spanish	0 IM: .5mL/w
Kel	he/they	26	Illinois	White	<i>n/a</i>	8 IM: .2mL/w (1,3-7,9-12), .5 (2); off HRT (8)
Kevin	he/they	38	California	Hispanic	Spanish	2 Gel: 50mg/d; SubQ: .25mL/bw (9), .25 (10), .3 (11-12)
Lizard	he/him	20	Illinois	White	<i>n/a</i>	4 SubQ: .3mL/w
Marshmallow	he/him	30	Texas	White	Spanish	3 IM: .3mL/w (1-3,6), .4 (4-5)
Oz	they/them	19	Arizona	Filipino, White	<i>n/a</i>	15 IM: .5mL/bw
Panda	he/they	18	Illinois	Ashkenazi, African American	<i>n/a</i>	0 SubQ: .3mL/w (1-9), .25 (10)
Raine White	he/him	27	New Brunswick	White	French	2 SubQ: .5mL/w (1-5), .7 (6-11), .6 (12)
Renaud	he/they	27	British Columbia	Mixed	French	5 IM: .25mL/w (1-4), .35 (4), .5 (6-11), .6 (12)
Saint	he/they	24	California	Latine	<i>n/a</i>	2 IM or SubQ: .25mL/bw (1), .5 (2-12)
Southern Boy	he/they	21	Louisiana	White	<i>n/a</i>	3 SubQ: .3mL/w (2-5), .4 (6-9), .5 (10-12)
Ty	he/him	29	Indiana	White	<i>n/a</i>	10 Gel: 50mg/d (1-5,8-10,12), 37.5 (6-7,9)
Wizard	he/they	21	Alabama	White	<i>n/a</i>	1 SubQ: .3mL/w (1-3), .4 (4-8), .5 (9-12)

Table 3.1: Summary of participant demographic information, as provided. Includes alias, pronouns, age, location, languages spoken, number of week on HRT as of the first session (S1), and HRT dose. HRT application may be subcutaneous (SubQ), intramuscular (IM), or via topical gel (Gel); dose may be daily (d), weekly (w), or biweekly (bw) and is followed by the weeks during which it was prescribed.

Greater emphasis in the past few decades has been placed on viewing trans issues through an intersectional lens, predominantly in the field of sociology (Schilt and Lagos, 2017). Sociolinguistics must do the same. Work by Bucholtz and Hall (2004, 2005) has provided frameworks for understanding both gender and sexual identity as the *product* rather than a *source* of their linguistic correlations; as said in §2, it is less that "male" speech is characteristic of men, and more that it is used to index maleness. Maleness and masculinity are defined differently across individuals and across cultures, and so it is impossible to examine gender and language isolated in a sociocultural vacuum. To that end, there are several specific elements of the participant demographics which I would like to highlight, not only because of how they affect the degree that this study is representative of the North American trans population, but because they are salient to speakers, and should thus be considered closely in our discussion of gender and the voice. These include age, race, and neurodivergence.

3.3.1 Demographic highlight: Age

Speakers ranged in age from 18 (Panda) to 53 (Gareth), including six in their 30s or older. In a survey of 421 transgender individuals seeking hormone therapy in a particular clinic, Leinung and Joseph (2020) found the mean age of FTM patients to be 26.2 ± 8.6 at the start of transition; the average age of speakers of this study is 28.2. The inclusion of trans speakers in their thirties and fifties is of importance here; as discussed in conversations with those participants, online community spaces tend to skew younger, and trans people outside of their teens and early twenties can find it difficult to find others with whom to relate.

The 2015 US Transgender Survey (UTSC) (James et al., 2016) shows a generational divide in identification as trans men or women versus non-binary, with 47% of binary individuals being between 25 and 44, compared to 35% of non-binary responders; in short, non-binary individuals were more likely to be younger. This is unsurprising; recent generations are undergoing self-discovery in a time when non-binary people are more open, out,

and recognized than ever before (Schilt and Lagos, 2017). This is not to say, however, that trans individuals out of their twenties are rarely non-binary; indeed, by the end of the current study, we find that three of the six speakers over 30 (Arrow, Gareth, and Kevin) identify as such (see §3.4.2. We additionally see, however, that the three speakers in the 30+ cohort¹ who did *not* do so did not necessarily consider themselves *binary*. Chimmy, Freddy, and Marshmallow all indicated that they would not use either *binary* or *non-binary* to refer to themselves, an answer they shared only with Lizard (who, at 20, is far from this cohort). In effect, these speakers defy a second binary: that of the (fluid, arbitrary, and ultimately self-defined) distinction between *binary* and *non-binary*.

Whether or not this perspective on binary versus non-binary identity is in fact a function of being (save for Lizard) the study's oldest, age does require consideration in the contexts of satisfaction for three reasons: because there is a cultural conflation of trans identity with youth, because transitioning affects one's social situation differently after having already developed a career, family, or both, and finally because there are some physical concerns regarding the effects of HRT as one ages. Online trans community spaces now thrive on social media that is most accessible to generations of digital natives. In-person community spaces are often centered around colleges and universities. Discriminatory perspectives and laws targeting trans people focus on youth: at best, trans people's existence is viewed as a fad spreading among teens and twenty-somethings; at worst, trans children and teens are held as examples of some corruptive power of "gender ideology", with multiple legal rulings in the United States during the current year focused on restricting the ability for minors to transition. From multiple corners of society, we see an association of transness with youth.

"[E]very time I went to a queer meetup full of college students or saw someone tweet about their "late transition" at age 25, I [...] aged from egg to elder² in an instant, like

1. I recognize the fallacy of grouping speakers into 18-29 and 30+ groups; I do so here only as a convenience in discussing age, and because the vast majority of participants of the current study are in their 20s.

2. Doyle, 40 at time of writing, is referencing here the ideas of "eggs" (individuals unaware or only

a thousand-year-old mummy being exposed to sunlight and crumbling to dust.

(from *The unique challenges and rewards of transitioning later in life* (Doyle, 2022))

Transitioning after one has been in the workforce for some time, developed strong community bonds, or started a family can also present unique challenges, and all of these things grow more common as one ages. Several participants in this study are married, and several have children. These speakers frequently related the trials and tribulations of parenting while trans, from the euphoria of being called "Papa" to judgement received from other parents. Experiences with their changing voices were often framed by this context; between his tenth and eleventh sessions, Chimmy began vocal therapy, and he described the goals he'd outlined during the initial appointment as follows:

I talked about not wanting to be consistently ma'amed on the phone, um, and wanting to find enough volume that I could get my kids' attention in a crowded room. [...] A couple of times, I've been in a crowd with my kids, and I haven't been able to project enough for them to hear me, and that's a legitimate safety issue.

(Chimmy S11|W60)

After misgendering, a practical concern regarding his children was his primary concern. Finally, some participants had concerns regarding the physiological effects of HRT, and how they may be affected by age. There is evidence that starting HRT as a teen leads to better mental health outcomes (Turban et al., 2022) and conversely that older (60+) trans adults tend towards better quality of life than younger ones (Cai et al., 2018). In terms of vocal change on HRT, current data is largely inconclusive, but there is substantial folk knowledge in online spaces expressing the general sentiment that the younger one is upon starting HRT, the better one's results (Ty S11|W62).

newly having realized that they are trans) versus "trans elders", individuals who have been transitioning (or transitioned, depending on to what extent one feels the process can ever be complete) for multiple years. Though often used in reference to trans people who are actual elders in the sense of age ($\approx 60+$ years), it can also refer to any such person who acts as a mentor and sources of advice to younger trans people earlier in the transition process.

I'd been online and heard people talk about, like, after 25 or whatever, your vocal chords have a harder time. [...] It might've been misinformation, but I'd seen it enough times that I was like, well, I'm older than 25, so I might have some issues with that.

(Ty S11|W62)

3.3.2 *Demographic highlight: Race*

Much prior work with trans speakers either does not report speaker race, or has involved a White, racially homogenous cohort (de Vries and Sojka, 2020)³. White speakers represented 62% of responders to the 2015 US Transgender Survey, with the next largest group being Latino/a (16.6%) or Black (12.6%). However, 31/34 (91%) of Papp (2012)'s total participants were White, as were 13/15 (87%) of Zimman (2012)'s. In the current study, 12/18 (67%) of participants self-identify solely as White, Arrow as Black, Kevin as Hispanic, Oz as Filipino and white, Panda as Ashkenazi and African American, Renaud as mixed, and Saint as Latine⁴. This is thus a slightly unusually racially heterogenous group of speakers.

Within the United States we have seen the evolution, and recent documentation, of language varieties and patterns of change associated with particular racial and ethnic groups. For example, while American linguists have long described variation in African American Language (AAL), modern efforts focus not only on its phonetic and grammatical attributes but on how it relates to segregation, place and identity, and the culture and history of its speakers (Farrington et al., 2021). Work with Latino speakers of Spanish in the US similarly considers place and identity, bilingualism, and the ways in which the development of uniquely American varieties of Spanish are as influenced by social and socioeconomic factors as articulatory ones (Lipski, 2008).

3. Studies tend also to skew towards assumptions of a middle-class, urban experience (Abelson, 2016). While I did not ask participants' to divulge socioeconomic status, I will say that many were working-class, a few were middle-class, and several were college students with varying financial situations and upbringings.

4. A gender-neutral alternative to Latino/Latina.

For many, particularly LGBTQ people of color, race cannot be divested from gender. Race affects both how speakers of various genders differ in production (e.g. African American speakers not showing a gender distinction in production of /s/ (Calder and King, 2020) as was found in white speakers (Podesva and Van Hofwegen, 2015)) and in how they are perceived (e.g. an association of "Blackness" with "maleness" leading to greater perceived masculinity of African American women (Goff et al., 2008)). This was salient to several speakers, and mentioned as contributing to their ability to pass in different social contexts ("I've noticed different cultures will interpret me differently. [*Being read as male*] will never happen with African American men." - Arrow S1|W8). However, it was not a common topic in interviews, and I will largely leave discussion of individual participants' experiences of race and gender to a later, more thorough exploration of those conversations.

3.3.3 Demographic highlight: Neurodivergence

Of particular note is that over half⁵ of participants identified as *neurodivergent*, a broad term describing individuals with differences in neurological function and cognition that are typically viewed in a medical framework as neurodevelopmental disorders (contrasting with *neurotypical*) (Kapp, 2020). This can include attention deficit hyperactivity disorder (ADHD), autism spectrum disorder (ASD), specific learning disorders (e.g. dyscalculia, dyslexia, dysgraphia, or dyspraxia), intellectual disabilities (IDs), and a wide variety of other conditions. I use the term "identified as" here because the specific word *neurodivergent* is not a universal umbrella term for those with these conditions, but rather is associated with a specific framework of human brain function, and an accompanying social movement working for acceptance of those with non-normative cognition.

5. This was not included in the demographic questionnaire, but discovered through interviews. I refrain from giving the exact number here, or speaker-by-speaker in Table 3.1, for two reasons: first because at least one participant explicitly requested I not directly reference them as such, and also because multiple participants suspected but were not certain whether this term could be applied to them (i.e. were undiagnosed).

Though the specifics of this framework and movement are not relevant here⁶, it must be acknowledged that not all people with cognitive differences would use the term neurodivergent in reference to themselves. In this way, it is not unlike the complex history and web of terminology used within LGBTQ communities, up to and including surrounding the word *queer* itself (§1.1). Like *queer*, associated with *neurodivergent* as a term is the idea of a *neurodiverse community*, a collective of people with various neurological differences who work together for mutual advocacy. And also like *queer*, the term *neurodivergent* allows individuals to avoid specifying the exact nature of their neurological function; it thus acknowledges the overlap of many of the aforementioned conditions.

This is relevant for two reasons: First, there is substantial evidence that neurodivergent individuals are more likely than the neurotypical majority to also be transgender, and that this presents unique challenges in pursuing transition (Strang et al., 2018; Kallitsounaki and Williams, 2022). The high number of neurodivergent participants in this study is noteworthy on its own; outside the scope of this work, the overlap of neurodivergent and trans experiences is a necessary topic of research, as activism from both groups continues to be part of the modern zeitgeist. Second, multiple participants in this study discussed how neurodivergence affects their perceptions and experiences of gender, as well as their interactions with others. ATM, for example, described its effects on how he processes language; he notes accent mimicry and the relative ease with which he picks up other languages, as well as the significance of certain foreign languages and cultures as special interests⁷. Some also drew parallels between concepts such as *masking* (camouflaging or suppression of certain behaviors with the intent of being perceived as neurotypical) and *passing* as it is discussed

6. For a thorough overview, see Kapp (2020).

7. The term "special interest" refers specifically to the intense interests and passions of many people on the autism spectrum; they are distinguished by their intensity, with such individuals often prioritizing them both as activities to engage in and as topics in conversations with others. A trans speaker with a special interest in a language-related topic might thus be predicted to be highly aware of the characteristics, gendered and otherwise, of their own voice. This is exactly what we see with ATM.

in trans contexts. Among those for whom it is relevant, there is significant curiosity towards, as Gareth put it, "the Venning of trans and neurodiverse experience" (Gareth S3|W5).

3.4 The identity questionnaire

As discussed in §1.1, there is substantial variation in the terminology preferred by trans individuals to refer to themselves, the trans community, and the wider LGBTQ community as a whole; Zimman describes this diversity as speakers "us[ing] the linguistic resources available to them to construct cohesive gendered styles that reflect their complex and variable relationships to concepts like men, maleness, and masculinity" (Zimman, 2012, p2). Prior linguistic work with transmasculine speakers has frequently, though not universally, recruited participants under very specifically-worded identities (e.g. *trans men*, *female-to-male transsexuals*, *female-to-male gender dysphoric individuals*). While this wording allows for clear delineation of a study's subject pool, there is significant nuance to trans identities which makes any such group categorization difficult. The identity questionnaire thus allows for a better understanding of how the participants of this study refer to themselves.

3.4.1 Structure of the questionnaire

The questionnaire consists of a series of 36 terms for gender and sexual identity, selected through discussion with individuals on the same online spaces in which participants were recruited (§3.2). These are the most common terms that members of the polled communities preferred to refer to themselves, as well as several (*woman*, *girl*, and *female* in particular) that were nearly entirely dispreferred, included here for comparison. An initial list of 100 terms was whittled down based on consensus about which were the most common and appropriate terms, with additional suggestions given by community members. Though generally judged "no longer common", *transsexual* was additionally included to acknowledge its historical significance and as I expected older speakers may still prefer it as a term. Construction of

this list was done organically in order to create an accurate reflection of the current most-preferred terminology; LGBTQ spaces are a hotbed of neologisms and novel expressions, and as such any list grows swiftly out of date. As it is, this one is far from comprehensive.

Study participants were instructed, for each term, to indicate how likely they were to use that term to describe themselves (on a five-point Likert scale of *Very Likely - Likely - Neither Likely Nor Unlikely - Unlikely - Very Unlikely*); a write-in option was also provided, where participants could elaborate on their responses or add additional ones that were not in the provided list. Following the completion of the questionnaires, interviews occasionally focused on these questions of identity, as well speakers’ relationships with certain terminology and with community spaces that find solidarity in shared labels. The full list is given below.

man	trans man	masculine	queer
woman	trans boy	feminine	straight
binary	transgender	masc	gay
non-binary	transsex	fem/femme	mlm
male	transsexual	androgynous	bisexual
female	transmasculine	gender-conforming	pansexual
boy	FTM	gender non-conforming	asexual
girl	FTNB/FTNBi	genderqueer	male-attracted
trans	AFAB	genderfluid	female-attracted

Identity is a fluid, evolving concept. This is especially true for those in the first year of medical transition, who face tremendous physical and emotional changes, and for some of my participants, concurrent upheaval in lifestyle and in access to community support. While both gender identity and sexuality are typically assumed to be determined relatively early in life and stable throughout adulthood (Swaab, 2007), physical and social changes over the course of transition can affect one’s perspective of both.⁸ Even if a given term (e.g. *trans men*) is an entirely accurate descriptor of all participants at the beginning of the research, there is no guarantee that this will still be the case by the end of it.

8. For example, a gay trans man might recognize that he is attracted to men only after he realizes he is trans; he may have been unable to picture himself as a woman with a male partner, assuming that it was the partner’s gender (rather than his own) that was at issue.

Meier et al. (2013) found that as many as 40% of female-to-male transgender individuals out of a sample of 605 experienced a change in sexual orientation over the course of their transition, associated with the beginning of HRT. Auer et al. (2014) found up to 22.2% of (in their wording) female-to-male transsexual persons to report a change in attraction, with those attracted to men showing the greatest likelihood of change, and in contrast did not find any correlation with the start of HRT. Shift in identity is thus not uncommon over time on HRT; the consideration of speaker identity and personal vocal goals in the discussion of satisfaction (as described by Azul 2015a) should take this into account, and longitudinal work in particular should not rely on pre-study demographic surveys or the like to remain an accurate gauge of speaker identity throughout the study.

A quantitative analysis of the change over time seen in the identity survey's responses is not included in this work, which focuses primarily on how these responses inform participants' responses to the second (satisfaction) questionnaire (§6). However, a broad overview of speaker responses will serve to illustrate the variety of identities found in this small sample of the transmasculine population. Importantly, I must clarify that list and questionnaire are not intended to reduce participants' experiences to some quantifiable subset of labels. Labels *are* important to many; they can help define group membership and foster camaraderie, or on an individual level can allow one to put words to a concept one never before had the ability to express. They serve the purpose of allowing us to talk about how these identities intersect, and how individuals can embody various aspects of gender in identity and presentation and might outwardly appear to contradict, or that might change over time. But they are not neat little boxes; no two speakers' definitions of *man* or *queer* or *non-binary* are likely to be the same, being less informed by dictionary entries than they are by the context of everything else that these speakers are as people.

	F	L		F	L		F	L		F	L
<i>man</i>	11	11	<i>trans man</i>	15	14	<i>masculine</i>	15	15	<i>queer</i>	16	16
<i>woman</i>	2	0	<i>trans boy</i>	9	9	<i>feminine</i>	4	7	<i>straight</i>	1	2
<i>binary</i>	3	1	<i>transgender</i>	17	16	<i>masc</i>	15	15	<i>gay</i>	15	15
<i>non-binary</i>	11	11	<i>transsex</i>	3	2	<i>femme</i>	2	4	<i>MLM</i>	6	4
<i>male</i>	10	9	<i>transsexual</i>	5	2	<i>androgynous</i>	7	5	<i>bisexual</i>	9	7
<i>female</i>	1	1	<i>transmas.</i>	15	15	<i>gen. conf.</i>	1	2	<i>pansexual</i>	5	4
<i>boy</i>	10	12	<i>FTM</i>	12	16	<i>gen. n-conf.</i>	14	11	<i>asexual</i>	2	2
<i>girl</i>	1	0	<i>FTNB/FTNBi</i>	4	3	<i>genderqueer</i>	10	7	<i>male-attr.</i>	10	7
<i>trans</i>	17	16	<i>AFAB</i>	13	14	<i>genderfluid</i>	4	3	<i>female-attr.</i>	11	7

Table 3.2: Number of "(very) likely" responses in (F)irst and (L)ast sessions.

3.4.2 Results of the identity questionnaire

The terms most-used by participants are *trans* and *transgender*, followed by *queer*, and then by *trans man*, *transmasculine*, and *masc*. The overall number of participants overall who preferred particular terms stayed relatively consistent throughout the study (Table 3.2); however, these were not necessarily the same speakers throughout, and changes were seen that, while not indicative of an overall trend, are relevant to individuals' experiences. I will discuss these in brief here, focusing on final session responses as presented in Table 3.3.

Man, woman, boy, girl, male, female: Most participants did consider themselves men, with four (Arrow, Gareth, Lizard, and Panda) ambivalent to the term. While the number of speakers preferring this term stayed the same between first and last session, individual responses did change, with Ty and Chimmy moving to disagreement over time while Kevin and Wizard moved towards agreement. All those who at the beginning did to some extent identify with *woman* or *girl* (Kel and Lizard) did not by their final sessions (see Kel S11|W62 below); more speakers, however, come to consider themselves *feminine* or *femme*, with speakers such as Southern Boy and Wizard (who had initially "strongly disagreed" with calling themselves *femme*) later preferring the term. Approximately half of speakers considered themselves male; again, though the number is consistent, the speakers are not. Arrow and Saint moved from unlikelihood to ambivalence, Chimmy from likelihood to unlikelihood. Only Lizard consistently throughout the study was likely to use *female*.

Terms	Ar.	AT.	Ch.	El.	Fr.	Ga.	Ke.	Kv.	Li.	Ma.	Oz.	Pa.	Ra.	Re.	Sa.	So.	Ty.	Wi.
man	x	●●	○	●	●●	x	●●	●	x	●●	○○	x	●●	●●	●	●	○	●●
woman	○○	○○	○○	○○	○○	○○	○	○○	○○	○○	○○	○○	○○	○○	○○	○○	○○	○○
binary	○○	○○	○○	○○	○	○○	x	○	○	○	○○	○○	●●	x	○	○	○○	○○
non-binary	●●	●●	○○	●●	○○	●	●	●	○○	○○	●●	●●	○○	x	x	●	●	●●
male	x	●●	○	○○	●●	x	●	●	○	●●	○○	●	●●	●	x	x	●●	○○
female	○○	○○	○○	○○	○○	○○	○○	○○	●	○○	○○	○○	○○	○○	○	○○	○○	x
boy	●	●●	○○	●	●●	○	●	○○	○○	●	○○	●●	●	○	●	●	●●	●●
girl	○	○○	○○	○○	○○	○○	○	○○	○○	○○	○○	○○	○○	○○	○○	○○	○○	○○
trans	●●	●●	○	●●	●●	●●	●●	●●	○	●●	●●	●●	●●	●●	●●	●●	●●	●●
trans man	●	●●	○	●●	●●	●	●●	●●	○	●●	○○	●●	●●	●●	●●	●●	x	●●
trans boy	●	●●	○○	●	●●	○	○○	○○	○○	●	○○	●●	●	x	●●	x	○	●
transgender	●●	●●	○	●●	●●	●●	●●	●	○	●●	●●	●●	●●	●	●●	●●	●●	●●
transsex	x	○○	○○	○○	○○	○○	○	○	○○	●	○○	x	●	x	○○	x	○○	○○
transsexual	○	○○	○○	○○	○○	○	○	○	●	○	○	●	x	x	○○	x	○○	○○
transmasculine	●●	●●	○	●●	●	●●	●●	●	○○	●●	x	●●	●●	●●	●●	●●	●●	●●
FTM	●	●●	○	●●	●●	●●	●	●●	x	●●	●	●●	●●	●●	●●	●●	●●	●●
FTNB/FTNBi	●●	○○	○○	○	○○	x	○○	○	○○	○○	○	●	○○	x	●	x	○○	○
AFAB	●	●●	●	●	●●	●	●	●	○	●	○	●●	●●	x	x	●●	●	●
masculine	●	●●	○○	●●	●●	●	●●	●	○	●●	○○	●	●●	●●	●●	●	●	●●
feminine	●	○○	○○	x	○○	●	○○	○○	x	○○	○○	●	○	○○	○	●	●	●
masc	●	●●	○○	●●	●●	●●	●	●	○○	●●	○○	●	●●	x	●●	●	●●	●●
femme	●	○○	○○	●	○○	○	○	○○	○○	○○	○○	●	○○	○○	○	●	○○	○
androgynous	●●	○○	○○	●	○○	●	x	x	○○	○	○	●●	○	x	x	x	○	●
gen. conform.	○○	○○	○○	○	○	○○	○○	○○	○	○	○○	○	●	x	●	x	○○	○○
gen. n-conform.	●●	●●	○○	●●	○	●	●●	●●	○	○○	○	●	○○	x	●	●●	●	●●
genderqueer	●●	●●	○○	○	○○	●●	●●	x	○○	○○	○	○○	○○	x	●	●	○	●●
genderfluid	●●	○○	○○	x	○○	●●	x	○	○○	○○	○○	○○	○○	x	○○	x	○○	●●
queer	●●	●●	●●	●●	●●	●●	●●	●●	○○	●	○○	●●	●●	●●	●●	●●	●●	●●
straight	○○	○○	○○	○	○○	○○	○○	○	○○	○	○○	○○	○○	●	●	○○	○○	○○
gay	●	●●	○○	●●	●●	●●	●●	●	●	○○	●●	●●	●●	x	●●	●	●●	●●
MLM	○○	x	○○	○	○	x	●	x	○○	○	○○	○○	○○	○○	○○	●	x	○○
bisexual	●	○○	○○	●●	○○	○	●●	●	●●	●	x	●	○○	x	○	x	○○	○○
pansexual	x	●●	○	●	○○	x	○○	●●	○○	●	○○	x	○○	x	○	x	○	○○
asexual	○○	○○	○○	●	○○	○	○○	○	○○	○○	○○	○○	○○	x	x	x	●	○○
male-attracted	x	●	○○	x	○	●●	●	x	○○	●	○○	●●	●●	○○	○○	x	●●	○○
female-attracted	●●	●	○○	○	○○	x	●●	○	○	●●	○○	●	○○	●	●●	x	○○	○○

Table 3.3: Final-session responses to the identity questionnaire for all participants. Bullets indicate response strength: ○○ *very unlikely* | ○ *unlikely* | x *neither likely nor unlikely* | ● *likely* | ●● *very likely*.

Binary, non-binary: Here again we see substantial change; only three speakers, a remarkably low number, considered themselves *binary* in the first session (Freddy, Kevin, Raine White). By the final session, only Raine White did so; they discussed at length how entering and leaving different community spaces, and interacting with other trans people, aided in that self-discovery (Raine White S11|W65). Several speakers eschewed both *binary* and non-binary as labels (Chimmy, Freddy, Lizard, Marshmallow) or in the case of Renaud were entirely ambivalent about both throughout the year.

At first I thought I was non-binary. Just smack-dab in the middle, not really genderfluid, just kind of there. So I joined some non-binary groups on Facebook. And then, the more I understood myself, the more I pretty much just came to the conclusion that I'm a binary trans man, so I joined *those* kinds of groups [...] and it just really helped, seeing everyone else's experience. It's made all the difference in the world.

(Raine White S11|W65)

Trans, trans man, trans boy, transgender, transmasculine: Of the array of trans-prefixed terms, participants showed variation. *Trans* was preferred, though not by either Chimmy or Lizard, who also were the only two unlikely to use *transgender*, *transmasculine*, and along with Oz, *trans man*. Marshmallow and Raine White, conversely, were the only two speakers to use *transsex* by the final session. Another term not listed here, *transperson*, was adopted by Gareth partway through the study (Gareth S10|W47), illustrating how changing perspectives and identity can continue throughout the process of transition.

I had the very rare occasion of speaking to someone about my trans-ness [...] and I noticed myself hesitating, like, still working out conversationally if I want to describe myself. [...] I'm still trying on all these words and things. Like, how do they fit? And what came out the most natural [...] is *trans person*. And that is feeling like actually the best term so far, is *trans person*. I'm preferring it even to *trans masc*. I'm a trans person. What does that mean? Well – [*laugh*] stick around, maybe you'll find out!

(Gareth S10|W47)

The use of *man*, *boy*, *trans man*, and *trans boy* varied for many speakers throughout the year. Within the communities where this terminology list was sourced, *boy* was not infrequently used by transmasculine people, particularly those in their late teens and early twenties, who were uncomfortable calling themselves *men*⁹, or those of any age who simply preferred a term that brought to mind youth (and, potentially, a cis boyhood that one never got to experience), rejuvenation or reinvention, and having one's whole life ahead. Many speaker who did not use *boy* or *trans boy* cited age as the primary reason (e.g. "I'm not a boy, I'm a grown adult"), but of those who did, three (Arrow, Marshmallow, and Freddy) are over 30; thus it does not appear to be a strictly age-specific term.

It is noteworthy that more speakers identified with *trans man* than with *man* alone. Though the total number of "(very) likely" responses for *man* remained the same, the responders did not. Chimmy and Ty went from "likely" to "unlikely", Arrow, Gareth, and Lizard in the opposite direction, all becoming ambivalent. Lizard in particular described avoiding labels, but beginning to experiment more with calling himself a man over time both as he transitioned and as he began to socialize more in LGBTQ spaces (Lizard S8|W34).

Labels don't really – like, it's not something that defines my life. Like, at all. [...] I feel that there's never a label that describes exactly one thing, ever. So it's just not something that I live by, especially. Although I've been thinking a lot more about how I fit in with others. [...] And it almost feels a little uncomfortable? 'Cause, usually I would just – completely abstain from using any kind of descriptor at all. But like, I now feel like maybe it's a little better. Maybe I'm allowed to describe myself as a man? Maybe? Kind of testing that water. (Lizard S8|W34)

Transsex, transsexual: Gareth, Lizard, Marshmallow, Oz, and Raine White all indicated usage of *transsexual* in their first sessions, but only Lizard and Panda in their last;

9. Either simply due to age, because did not think they were "men" *enough* or deserving of the label pre- or during-transition, or due to negative experiences with men such that it was not a group they wanted to be a part of.

Raine White, Renaud, and Southern Boy were ambivalent. Similarly, only Marshmallow and Raine White used *transsex*¹⁰ Those who used these terms at various points expressed that they more accurately captured the medical component of their transition, and Lizard by Session 12 preferred *transsexual* to *transgender* as a better descriptor. This degree of use was somewhat surprising, particularly for *transsexual* given its reputation in the community as outdated; I had expected it to be relatively more common among older participants, but Panda and Lizard are in fact the youngest and third-youngest, respectively. While this is far too small a sample size to tell, I speculate that this term might be seeing a return in younger trans individuals, now several decades removed from its original pathologizing context.

FTM, FTNB/FTNBi, AFAB: Many participants clarified that they would use this terminology, but only in medical or technical contexts. Chimmy notably did not use *FTM* ("female-to-male"), and Lizard was ambivalent. *FTNBi* is rarer and nonbinary-specific¹¹, while the more clinical *AFAB* saw increased likelihood of use from Saint and Ty.

Masculine, feminine, masc, femme, androgynous: Speakers varied in their use of these terms, not only between first and last sessions but throughout the study. Most began and remained *masculine*, with Lizard becoming less so and Panda more so. More considered themselves *feminine* than *femme*, though rates of both increased. Several speakers (Arrow, Gareth, Panda, Southern Boy, Ty, Wizard) considered themselves both *masculine* and *feminine*; Chimmy and Oz notably considered themselves neither. These terms, depending on speaker, were used to describe both elements of internal identity and of external presentation.

Changes here typically came as speakers experimented with presentation, sometimes finding that differently gendered pronouns, names, clothing, and so on was only comfortable in the context of being perceived in different ways. For example, one might enjoy dresses, but

10. *Transsex*, as it was described to me by the communities where these terms were sourced, is a variant of *transsexual* intended to reduce its misinterpretation as a sexual orientation (à la *heterosexual*, *bisexual*, etc.) and to divest it from existing stigma.

11. The addition of the (i) is relatively recent, thus the inclusion of the previous version for clarity; it is used to disambiguate between the abbreviations NB (non-Black) and NBi (non-binary, formerly also NB).

only if in wearing them one is read as a feminine man, rather than as a woman. Alternatively, one may not consider such presentation *feminine* at all; one is simply a man (or person) in a dress. Testing boundaries of comfort in presentation and how one wishes to be seen is not a simple process, but one that continued for many speakers well into and beyond the first year of HRT. ("You try things and see what sticks, and then you have a floor that's covered in gender debris. All the bits. You're like, why did I buy this miniskirt?" -Kel S1|W8).

Notable here is the prevalence of speakers who consider themselves both *masculine* and *feminine*. As discussed in §2 and further in §6, there is a tendency to describe gender as a spectrum, but a linear one, such that change away from one end is necessarily also change towards the other. Reminiscent of the Kinsey scale's (Kinsey et al., 2016) presentation of hetero- and homosexuality as inversely related, male and female are seen as two ends of a sliding scale, with non-binary individuals somewhere in between. Similarly, masculinity and femininity are seemingly mutually opposed, with androgyny the space in the middle.

It becomes increasingly apparent, however, not only in examination of existing literature on trans voices (which may use any number of sociological models of gender) but *in conversation with participants* that, at least in this small sample of the wide trans population, this is not the common view. Rather, masculinity and femininity are framed to exist in parallel, as eloquently described by Gareth (S3|W5).

Having been so aware of my trans-ness for so long without transitioning, I have been – quite cognizant that it is precisely *through* the lens of masculinity that I embrace my femininity. And I don't see masculine and feminine as polarities. Um – I honor the extent to which it takes a lot of strength and – and courage, to experience femininity as a male quality. And so, that all just seems perfectly natural to me. It doesn't seem like a conflict of interest. It – you know, femininity is very much a part of masculinity, and part of that for me is – you know, my orientation as a gay man, and a queer man.

(Gareth S3|W5)

Gender (non-)conforming, genderqueer, genderfluid: By the last session, only Raine White and Saint considered themselves *gender-conforming*, though this did not mean that all others used the term *gender-nonconforming* (or GNC). Indeed Chimmy, Freddy, Lizard, Oz, and Raine White all used neither, while Saint used both. Approximately half of participants referred to themselves as *genderqueer*, far fewer (Arrow, Gareth, and Wizard) as *genderfluid*. Interestingly, while none who were likely *genderfluid* were also likely to use *male*, the same was not true for *genderqueer*.

Worth discussion also is that identity as *non-binary*, *genderqueer*, or *genderfluid* can itself play into questions of satisfaction. While the ways an androgynous *voice* interacts with satisfaction are addressed in §6.3.1, *identities* other than binary man or women are not easily navigated in modern American (or Canadian) society. While many non-binary and genderqueer participants found fulfillment in those identities, some also expressed the belief that having a more binary identity, more clear-cut physical or social dysphoria, or otherwise adhering more closely to the "expected trans narrative" (§1.1) would have made self-discovery, and later transition, easier (Ty S10|W56).

There's so many different flavors of [the trans man experience] and you know, you see some people's accounts and they're like, very, very straight men. And it's like, good for them, honestly, but I'm still very genderqueer. And I think – I think there's still a lot of internal tension with that. Um, and internal dysphoria. Because part of me wishes that I could just, like, pick one [*laugh*] and be that, and just, like, really lean into it. And I think the – the tension and the like, sitting on the fence and mixing elements of different things, is just part of who I am. (Ty S10|W56)

Queer: In the first session, every participant save for Lizard and Oz referred to themselves as *queer*; Oz was ambivalent, while Lizard was "very unlikely" to use this term; by the final session, both were "very unlikely". As discussed in §1, while *queer* has been largely reclaimed for individual use and in academic work with LGBTQ communities, not all (par-

ticularly those who have experienced its use as a slur) necessarily are comfortable in that reclamation, as is the case for Lizard. For those who do use it, it may be a cornerstone of their identities; apart from *AFAB* in medical context, queer was the only term used by Chimmy – as he put it, "in every sense of the word" (Chimmy S1|W7).

Sexualities and attraction: Only two speakers, Renaud and Saint, ultimately identified as *straight*; Saint notably also considered himself *gay*, while Renaud did not. Neither preferred *bi(sexual)* or *pan(sexual)*, so we see here a case of Saint exhibiting what would initially appear a contradictory sexuality: straight *and* gay, but not bi. Like male/female and masculine/feminine, the two are not necessarily opposed. The vast majority (15/18) of speakers called themselves *gay*; of these, seven also used *bi*, *pan*, or both. Only Marshmallow preferred the latter and did not also use *gay*. Elias and Ty, in addition to *gay* (and in the former's case, *bi* and *pan*), also used *asexual*.

Change in sexuality over time was noted by several participants, often involving either increased interest towards men, decreased interest in directly labeling one's sexuality, or both. Many veered away from terms like *MLM* and *male-attracted* over the course of the year, instead coming to prefer terms like *queer* or responding ambivalently towards sexuality terms (e.g. Renaud and Southern Boy, see Table 3.3). Others specifically talked about becoming more comfortable in their attraction over the course of the year (Kel S11|W62).

A year ago, I described myself completely as non-binary, genderqueer. Um, really didn't use any gendered terms at all, and I was using strictly *they/them* pronouns, and identified probably most with "butch lesbian" as a sexuality. And now, after going on T, that ended up shifting to the point where now I would describe myself as, like, "bisexual non-binary trans guy". [...] I'm actually interested in sleeping with men again – which is wild – but like, in a gay way. (Kel S11|W62)

Terms like *male-attracted* and *female-attracted*, intended to describe an individuals' sexual attraction without referencing their gender (and thus include as non-binary people) were

dispreferred by the end of the study. All who were *male-attracted*, except for Marshmallow, were also *gay*; all who were *female-attracted* were also *bi*, *pan*, or *straight*. *MLM* (men-loving-men) was used by a subset of speakers (Kel, Panda, Raine White, and Southern Boy) who also called themselves *gay*, and all but Panda were indeed *men*, with Panda ambivalent towards that term; however, many speakers who were men that loved men (or male-attracted) did not call themselves *MLM*.

These results are notable for not reflecting the overall demographics of the US trans population according to the UTSC, which measured its responders' self-identification rate of queer (21%), straight (15%), gay (16%), asexual (10%), bisexual (14%), and pansexual (18%) James et al. (2016). While this census did not only include transmasculine speakers, it is nonetheless unexpected that this sample should skew so heavily gay; I am uncertain why this would be, except that it may reflect the demographic makeup of the particular online communities where participants were recruited.

Participants were also provided with a space wherein they could mention other ways they would describe themselves, be it sexuality, gender, or both. Over the course of the year, this included *boi* (4 participants), *agender* (3), *fag/faggot* (3), *androgynous* (2), *trans guy* (2), *butch* (2), *tomboy* (2), *femboy* (2), *demiboy* (1), *demisexual* (1), *autosexual* (1), *transneutral* (1), *enby* (1), *t4t* (1), *top* (1), *versatile* (1), *dyke* (1), *dom* (1), *masculine-of-center* (1), *nonbinary-attracted* (1), *polyamorous* (1), *autigender* (1), *bigender* (1), *multigender* (1), *polygender* (1), *genderfuck* (1), *genderflux* (1), *cassflux* (1), and *sapphic* (1). In the interviews, participants often elaborated on these terms, how their relationships with them changed over time, relating stories of how decades ago they stumbled across a word that fit and a community that welcomed them, or how just last week they found a word online that perfectly captured the intricacy of their personal relationship to gender and sexuality. As said, each participants' individual experiences with identity over the course of this year could themselves be a case study, and I fully intend to give this the attention deserved in future work.

3.5 Final considerations

It must be acknowledged how my own identity as a white, mid-20s trans person played a role in these interviews, and in how participants responded during them. Many mentioned that they would not have considered the study had the recruitment materials not explicitly mentioned that the interviewer, myself, was trans – information which I included in those materials because of similar responses described by Zimman (2012), and which multiple speakers within the first two sessions also confirmed by asking me.

Again following Zimman, interviews were frequently an exchange, with a casual back-and-forth conversational style; though I encouraged participants to do the majority of the speaking, I also when asked shared details or anecdotes about my own transition. Discussions at times grew personal, involving issues of sexuality, mental health, and similarly sensitive topics with which I likely would not have been entrusted were it not for my status as trans, and for my own openness about similar topics when asked. That said, membership in the same community as one's participants does not infer wholly shared experiences; the circumstances of my upbringing were extremely different from those of my speakers, and we differed in gender identity and sexuality as much as they differed from each other.

My insider status is not a wholly positive one; namely, it may increase susceptibility to bias in interpreting study results. It also does not eliminate the distrust present towards research in trans communities, born of fear that our experiences may be misrepresented or used disingenuously to support discriminatory legislation; even after developing these relationships with participants, there was still caution shown in what was shared, particularly in how negative experiences were worded. Many expressed excitement for the study and enthusiasm that this work was being done, with the (hopeful) eventual result of providing not only better understanding about how and why many transmasculine people grow happy with their voices after time on HRT, but potentially useful advice for those who do not. All who did so then also expressed relief that the work was being done by a trans researcher.

It is for that reason alone that I disclose my trans status in this work. My litany of concerns about doing so are too numerous to list, and so I direct one final time towards Zimman (2012, p91-93), who explains his own reasoning, which I share, more eloquently than I ever could. At the time of writing, much of the world remains hostile to trans individuals. Disclosure, particularly in a permanent record as this work will be, is not without risk. And yet, again and again among these speakers, in discussing both this study and (during the final-session debrief) other work by which it is inspired, I saw the appreciation that – as is becoming increasingly expected in other linguistic research – those of the group being studied are spearheading (or at the very least deeply involved with) linguistic work in their communities. Not only does disclosing my own identity provide necessary context for the dynamic in this study between interviewer and interviewee, but it is my hope that doing so can be a source of reassurance: many of us who study trans voices do also speak with them. Now if I may be permitted a moment of colloquialism, that is quite enough about me.

Overall, this participant cohort displays extensive variation, including in age, race, and neurodiversity as well as in gender and sexuality identity. The latter in particular sees change over time, as the experience of transition recontextualizes individuals' feelings towards their own bodies and towards others. We see an overall, but not universal, preference for speakers to call themselves *queer*, *trans*, and *gay*; we also see numerous instances of simultaneous use of terms that appear contradictory if viewed only on a linear scale (*male* and *female*, *straight* and *gay*, etc.) We must take this variety, this change, and this non-linearity all into consideration going forward, first into acoustic analyses of these speakers' pitches and vowel spaces, and then into discussion of satisfaction.

CHAPTER 4

PITCH AND HRT

The following chapter examines pitch (quantified by F0) across all participants, including change over time in habitual speaking pitch, pitch range, ceiling, floor, and the distance between habitual pitch and floor. It further discusses the mechanics of measuring pitch in trans speakers (in particular critiquing the use of read-passage tasks), patterns seen in rate of pitch change, and the impact of pre-HRT vocal training.

4.1 Methodology

As discussed in Chapter 2, an overall lowering of pitch is the most well-attested effect of HRT on the voice, with most individuals reaching an approximately cis-male-normative frequency ($F_0 \leq 131\text{Hz}$; (Gorton et al., 2005; Evans et al., 2008; Fitch and Giedd, 1999; Irwig et al., 2017)) while others (as many as 21%; (Ziegler et al., 2018)) reach a "gender-ambiguous" range ($F_0 \leq 185\text{Hz}$; Gelfer and Mordaunt 2018); though largely arbitrary, these pitch targets provide useful points of comparison. All participants who underwent HRT throughout the duration of the study exhibited this characteristic drop in pitch. This includes a lowering both of overall pitch range, calculated from the minimum and maximum comfortably sustained pitches for each participant, and of mean habitual speaking pitch, obtained by taking the mean F0 of relevant vowel tokens extracted from a pair of elicitation tasks.

Historically, work with trans voices has focused on eliciting audio by asking speakers to read aloud a short text. This is often the *Rainbow Passage* (Fairbanks, 1960, see Appendix), which is widespread in phonetics research broadly and work with trans speakers specifically (Zimman, 2012; Ziegler et al., 2018; Hope and Bradley, 2019, among others). It contains a wide variety of environments for relevant phonemes, allowing the collection of numerous tokens for analysis, and allows comparison with other studies that use these same tokens.

However, there are several potential complications with this method. First, there is evidence that the acoustic properties of speech elicited via reading aloud differ significantly from those found in naturalistic conversation, including in fundamental frequency for women (Zraick et al., 2000; Schiwitz, 2011), spectral space (Nakamura et al., 2008), and intonation (Howell and Kadi-Hanifi, 1991). As such, we may question how representative of conversation read speech truly is, both in a broader linguistic context and in trans-specific work. We must also consider the prevalence of reading aloud as part of vocal training for trans speakers, either under the guidance of a vocal coach, individually, or through community support groups. Multiple participants reported recording themselves reading aloud, both pre-HRT and during the study, as a means of practicing voice masculinization techniques; indeed, so ubiquitous is the *Rainbow Passage* that several participants were already familiar with it in that context, and clinical guides for trans-specific vocal therapy explicitly recommend its use (Gelfer and Mordaunt, 2018, p193).

For trans speakers, reading aloud allows for focus and fine-tuning of control, potentially producing a more optimal voice result; one can focus on the sound of one's speech, rather than its content. It must be recognized that trans people interacting in the day-to-day world, speaking with strangers or in public, are *not* reading aloud, and cannot rely on the careful performativity allowed by read speech. Indeed, while we may expect read speech to reflect participants' desires towards their own voices, it may not reflect how those voices are perceived by others in spontaneous, unrehearsed contexts. This is a concern echoed by study participants, who cited on occasion the perception that when reading a passage aloud at home, slowly and carefully, they were more aware of their voice and thus consciously making more effort to lower their pitch. This raises the possibility that the results of any prior work with trans speakers making exclusive use of read speech are in fact representative only of those speakers' *optimal* voice patterns, in a rehearsed and comfortable environment. This is not a failing; indeed, it suggests that read speech may be more reflective of that

speaker’s vocal ideals.¹ However, in the context of satisfaction specifically, we must contrast this with the unrehearsed utterances of speakers’ daily lives.

It may be the case that the differences between read and spontaneous speech are negligible in their effects on perceived vocal gender, or satisfaction more broadly. That said, in acknowledgement of the potential differences (particularly in F0 and vowel space), and of the salience and potential markedness of reading aloud in trans spaces, the read passage task is contrasted with one intended to elicit more spontaneous speech. This second task is an image description, in which participants were presented with a variation on the *Cookie Theft* image (Goodglass and Kaplan, 1983) and instructed to explain the goings-on of the scene. Habitual pitches measured from the reading and image description tasks are referred to as Mean Reading Pitch (MRP) and Mean Description Pitch (MDP) respectively.

4.1.1 *Methods and expectations: Read passage*

A number of word tokens found in the *Rainbow Passage* were selected for analysis from the vowel sets FLEECE /i/, KIT /ɪ/, DRESS /ɛ/, FACE /ej/, TRAP /æ/, STRUT /ə/, PRICE /aj/, GOAT /ow/, LOT /ɑ/, GOOSE /u/, and FOOT /ʊ/. Token selection prioritized vowels in the following order: *stressed* > *CVC* > *inter-obstruent*. Most vowels produced tokens conforming to all three restrictions, with the exceptions of high back vowels FOOT and GOOSE (the former found only in open syllables and the latter only following approximants [l] and [w], save for the token "super-imposition").

Keeping to these constraints where possible provides 43 unique tokens for MRP (Table 4.1), of which several appear multiple times. Tokens were identified in textgrid-annotated recordings produced by the Montreal Forced Aligner and manually checked for accuracy; analysis was done using Praat (Boersma and Weenink, 2023), and Christian DiCanio’s scripts

1. Zimman explicitly acknowledges this: "...instead of trying to use read speech to figure out how my participants talk when they aren’t reading, my interest is in read speech as a genre in its own right that can potentially reveal speakers’ desires and stances with respect to their own voices." (Zimman, 2012, p70)

Vowel Set	MRP Lexical Items
FLEECE /i/	people, Hebrews
KIT /ɪ/	physical, difference, give, since, mixed
DRESS /ɛ/	centuries, accepted, many, heavy, depends, said
FACE /eɪ/	take, shape, mainly
TRAP /æ/	path, passed, imagine, band
STRUT /ə/	sunlight, colors, above, others, number
PRICE /aɪ/	finds, sign, size, type
GOAT /oʊ/	token, bows, bow
LOT /ɑ/	pot, gods, not, upon
GOOSE /u/	superimposition, two, blue
FOOT /ʊ/	look, looks, looking, would

Table 4.1: Vowel tokens analyzed from the *Rainbow Passage* task

were used to extract each token’s fundamental frequency at the vowel’s midpoint. F0 was not always measurable for each token due to the presence of creaky voice disrupting both automated and manual pitch measurement; this was particularly the case for participants Oz and Arrow, who exhibited high levels of creak throughout the study. Where it was possible for Praat to identify and extract pitch data from creaky-voiced vowels, this was done. The potential lowering effect of creak on perceived pitch must not go unrecognized, but an analysis of speakers’ use of creak is beyond the scope of the current study.

Regarding analysis of pitch, as described in §2.1.1, Ziegler et al. (2018) referenced two target pitches in determining whether speakers achieved "successful" voice masculinization: 131Hz ("cis-male-typical") and 185Hz ("gender ambiguous"). These values are inherently arbitrary; we may just as easily consider 160Hz as a high boundary for perception of a voice as "male" (Spencer, 1988), and \approx 170Hz as a low boundary for gender ambiguity (Gelfer and Schofield, 2000), both based on transfeminine experiences. They do, however, provide a useful baseline for comparison with prior work and for forming predictions of speaker pitch; for 131Hz and 185Hz, Ziegler found "failure" rates of 20% and 0% respectively, defined as study participants not exhibiting mean speaking pitches at or below these F0s (as measured largely by read-passage tasks). As such, we may predict a similar distribution here, and will take these targets into consideration.

Finally, there are expected differences in the rate and amount of change during different periods of the first year of HRT. Prior work has consistently found the most significant drop in habitual speaking pitch to during the first 2–5 months (Deuster et al., 2016; Nygren et al., 2009; Irwig et al., 2017). While Deuster et al. shows no significant decrease in F0 between 12 and 24 months, Irwig et al. identifies several different patterns in F0 lowering among their participants: lowering either only or minimally in months 0-3, only in months 0-6, only in months 3-9, and consistently between months 0-12.

We thus predict that: **(1)** Mean habitual reading pitch (MRP) will lower with time on HRT, with all speakers reaching $F0 \leq 185\text{Hz}$ and approximately 80% reaching $F0 \leq 131\text{Hz}$, and that **(2)** Given the behavioral as well as biological influences on pitch, MRP will be lower in participants who report intentional efforts to lower their voice. Finally, **(3)** participants will exhibit different patterns of voice lowering rate throughout the first twelve months.

4.1.2 *Methods and expectations: Image description*

The image description task made use of a partially-colored version of Goodglass and Kaplan's (Figure 4.1) (1983). Colors were added to encourage the production of the tokens *color(s)* and *blue*, also appearing in the *Rainbow Passage*. Participants were presented with the image and asked very broadly to "Describe what is happening in [it], going into as much detail as [they] like".

This task allows for some control over the content of the elicited tokens (e.g. participants tended to use words such as *kitchen* and *cookie*) while encouraging more naturalistic speech. Though speakers over the course of the study did come to anticipate the task, and thus knew what they would be commenting on, their responses were largely unrehearsed and spontaneous. Importantly, participants were not explicitly told that the recording of this task would be used for acoustic analysis; though many would come to this conclusion by the final session (see §4.4.2), this was done initially to encourage less self-monitoring of speech.

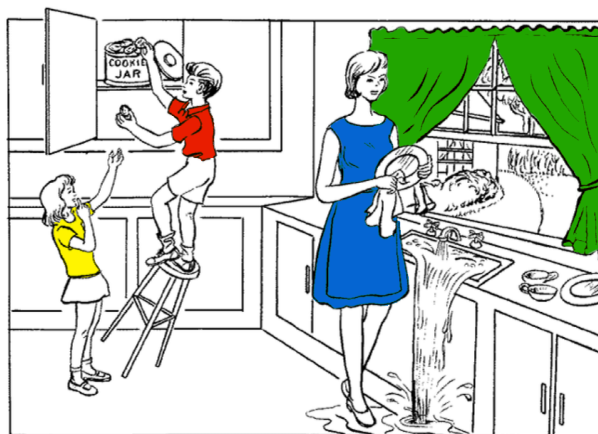


Figure 4.1: The colorized *Cookie Theft* image.

While certain words appeared frequently, participants were left to their own devices in word choice, creating no guarantee that viable tokens would be produced for every vowel in the preferred context (stressed, CVC, inter-obstruent) during each session. Wherever possible, at least one token of each vowel was selected per participant, per session; as such, the list of possible tokens for this task is much broader (Table 4.2). Italicized tokens are those which violate one of the indicated constraints, added as otherwise some participant would have no viable token for that vowel for some session. The mean count per session was 32 tokens, and a range of 25-35 was typical, though participants varied in verbosity from 79 for a given vowel (Wizard S7|W31) to 1 (Oz S9|W59).

Though most studies on which Ziegler et al.'s 185/131Hz targets were based did not use an image description task, we may still expect the same proportion of participants to exhibit voice lowering to each of these points. Thus, as with MRP, we expect that **(1)** MDP for all participants will lower with time on HRT, with all speakers reaching $F0 \leq 185\text{Hz}$ and approximately 80% reaching $F0 \leq 131\text{Hz}$, and that **(2)** MRP will be lower for participants who report intentional lowering of their voices. We also predict that **(3)** There will be significant differences in MDP and MRP across speakers, with the former being higher on average due to cited differences in read versus spontaneous speech production.

Vowel	MDP Lexical Items
FLEECE	beneath, deep, detail, details, features, feet, mean, meet, neatly, obscenely, people, piece, receive, scene, secretive, seem, seemed, seems, seemingly, serene, sneak, sneakers, speak, teacup, teetering, teeth, thief, underneath, <i>cleaning, completely, extremely, feel, feels, feeling, green, reach, reaching, stealing</i>
KIT	addition, dinner, dish, dishes, fifties, figure, kid, kids, kitchen, picture, sibling, sink, sister, tip, tipping
DRESS	again, attempt, attempting, attention, benefit, better, center, century, connect, depending, domestic, except, exception, extend, feminine, head, heads, hectic, gesturing, gender, gendered, get, getting, guess, guessing, men, mess, next, pandemic, perspective, presenting, said, sense, set, shed, step, test, together, <i>dress, neglecting, red, shelf, shelves, tell, wet, yellow</i>
FACE	basically, change, daydreaming, days, dazed, face, facing, favorite, gaze, gazing, improvisation, insane, jane, janes, making, maybe, neighbor, neighborhood, neighboring, neighbor's, pain, pane, panes, nature, same, shaded, shades, shading, space, stable, stage, state, taken, taking, <i>away, drapes, plate, plates</i>
TRAP	advantage, back, background, backwards, backyard, cabinet, cabinets, cat, disaster, fabric, fact, family, family's, fashion, hand, hands, imagine, manage, managing, manicured, masculine, matched, matching, panel, passes, passing, passive, past, path, pathway, saddle, Saturday, shadow, snacks, snatch, standing, stands, understand, <i>black, distracted, laughing</i>
STRUT	above, another, assumption, buckle, butt, coming, couple, covering, cups, cupboard, fun, fuzziness, jump, jumping, month, mother, mother's, number, puddle, pumps, shutters, skunk, something, son, stuff, suburbs, suds, touches, touching, tumble, <i>adult, brother, brother's, color, colored, colors</i>
PRICE	behind, design, dividing, driveway, find, fine, five, inside, kind, knife, side, sidewalk, ties, time, type, nineteen, outside, upside, <i>child, climbing, describe, dry, high, highlights, primary, smiling, white</i>
GOAT	associated, boasts, coded, combed, dissociating, focus, focusing, moment, mostly, motion, nose, notice, noticed, noticing, photo, shows, soak, stove, supposed, supposedly, zoned, <i>flow, know, overflows, overflowing</i>
LOT	accomplice, atop, bob, bothered, bottom, comic, commentary, cops, forgot, horizontal, jar, possibly, pompadour, probably, modern, mom, mom's, posture, socks, spots, spotty, stop, top, topple, toppling, <i>falling, washing</i>
FOOT	bush, bushes, cookie, cookies, good, pushing
GOOSE	assume, assuming, choose, moving, presume, presumably, shoes, <i>blue, cute, future, stool, tool</i>

Table 4.2: Vowel tokens analyzed from the *Cookie Theft* task

4.1.3 Methods and expectations: Pitch range

As described in §2.1.3, it is currently inconclusive whether we can expect a significant narrowing of pitch range during the first year of HRT. An accurate measurement of range is inaccessible through either the read passage or image description, as participants are unlikely to reach both pitch floor and ceiling over the course of either task. Thus, following the methodology of Papp (2012), at the end of each interview participants were walked through a brief range exercise. Those experienced with vocal warm-up techniques were instructed

to sing scales using the syllables /la/ and /mi/, while others were instructed to imitate a siren or fake a yawn; the method used remained consistent for each participant between sessions.² The highest and lowest notes comfortably sustained ($\geq 200\text{ms}$) by participants had their average F0 measured across their duration. Additionally, we will account for the argument that Hz is an insufficient metric of change in pitch range due to not reflecting the psychoacoustic *perception* of pitch, which is of particular concern here for its relevance to satisfaction. As such, we make use of the logarithmic semitone scale (ST) for measurement of both range (ceiling-floor distance) and mean-floor distance.

The choice of semitones instead of other psychoacoustic scales (e.g. Bark, Mel, ERB) is made for two reasons, aside from ability to compare with other studies that have used the same metric. First, among these scales, only the ST scale has remained unchanged for centuries (Papp, 2012, p34). Second, due to their prevalence in music theory, semitones are a salient method of discussing the voice, and one with which participants were often familiar; those with singing experience or other musical training would occasionally reference their changing vocal range using semitones, and so measuring them directly allows for comparison with participants’ own reported experiences. The formula by which semitones are calculated from the distance between floor and some pitch n is as follows:

$$ST = 12 \cdot \log_2 \frac{f_n}{f_{\text{FLOOR}}}$$

We also focus specifically on pitch floor, and how it compares to participants’ habitual pitches. Papp describes participants’ final habitual pitches as 6-13ST above floor, contrasting with the expected 3-4ST, and we may predict the same pattern here. We can relate this to the two broad patterns seen in how speakers adjust to a lowered floor: First, mean

2. Note that this work does not address other variables affecting range such as hydration (Alves et al., 2019), any vocal warm-up (Amir et al., 2005), and time of day, which could not be directly controlled for during the study, though the latter remained relatively constant for each participant due to interviews typically being scheduled at regular hours.

and floor may lower in parallel, with speakers maintaining the same comfortable distance from their floor while speaking. Alternatively, we may see pitch floor drop sharply, while mean pitch descends gradually; such participants become physiologically capable of reaching lower pitches, but do not immediately adopt this in speech. This "buoyancy" of habitual pitch is behavioral, due either to complexities of the "fine-tuning" process around gender presentation (Gelfer and Mordaunt, 2018), to a higher pitch being "more physically familiar, and more psychologically comfortable" (Block, 2017), or some combination of the two.

One possible factor in this buoyancy is whether participants use vocal training techniques to affect a lower speaking pitch pre-HRT, and thus intentionally speak close to their floor. Speakers reported by Papp to show increased mean-floor distance (MFD) described discomfort with the physical sensations associated with "chest voice", the use of which is a vocal masculinization technique. As such, we may predict speakers who use such techniques to be more comfortable with that feeling of resonance, begin with a smaller MFD, and exhibit the "parallel" pattern as they more quickly adapt to a lower floor. Furthermore, Papp found in some speakers an *increase* in mean pitch, not reflected in pitch floor and not explainable by the physical effects of testosterone (a "J" pattern), following speakers' first year of HRT.

Our predictions are thus: **(1)** Pitch floor and ceiling will lower over time on HRT; **(2)** Speakers will vary in the narrowing of their vocal range; **(3)** MFD will increase overall as speakers' habitual pitch adjusts more slowly than pitch floor; **(4)** Speakers who report speaking at an intentionally low pitch will show a smaller overall MFD; **(5)** If there are significant task effects in habitual pitch mean (read passage versus image description), there will also be a significant difference between MRP- and MDP-floor distances.

4.2 Initial-session pitch and pre-HRT voice lowering efforts

Examining first-session pitch results provides us with a baseline to which we may compare further change, and a focus on individual speakers; during these sessions, participants show

substantial variation in mean reading and description habitual pitches, as well as in initial pitch range. Habitual pitch aggregated across tasks ranged from 113Hz (Oz) to 228Hz (Saint). Pitch ceiling ranged from 375Hz (Arrow) to 733Hz (both ATM and Freddy), with a mean of 538Hz. Floor ranged from 96Hz (Oz) to 226Hz (Saint), with a mean of 153Hz.

During first-session recordings, participants were anywhere between pre-HRT (Elias, Gareth, Panda) and 15 weeks into HRT (Oz), with a mean of 4.1 weeks; this initial low F0 for Oz in particular may thus have been due in part to prior effects of HRT. Of those who were within one month of starting HRT, the lowest initial floor was that of Gareth (109Hz). Gareth had the third-lowest initial MRP of all participants (166Hz), second only to Oz and Ty (Table 4.3); Ty having been 10 weeks on HRT at the start of the project, second-longest after Oz, it is again likely that prior time on HRT is the cause of these initial low F0s.³

The lowest initial MRP was also Oz (110Hz); throughout the rest of this work, the late starting point for Oz continues to result in some patterns of change that may not be comparable to other speakers. A similar issue, though not evident in the first session, is that of Marshmallow, who completed only 7/12 sessions and was 41 weeks on HRT at cessation; comparison of Marshmallow’s final session with those of other speakers thus results in a similar issue to Oz’s own first session. To avoid the reiteration of these caveats whenever either speakers’ results are discussed, I will mark such results with ([†]); this will also be used in additional situations where further context is necessary to interpret a speaker’s data for some task and session.

Marshmallow[†] (162Hz) exhibited the lowest initial MRP among participants within one month of HRT; the highest initial MRPs and MDPs were 226Hz and 230Hz respectively, both from Saint. Speakers spoke with an average mean-floor distance (MFD) of 3.76ST; i.e. their habitual speaking pitch was approximately 3.76 semitones higher than their floor. Saint

3. Notably, Gareth’s initial F0 pre-HRT was comparable to those who were several months on HRT already. This may be attributed in part to age; Gareth is the oldest participant at 53, and habitual pitch across genders is known to lower with age, particularly for those who have gone through menopause (Eichhorn et al., 2018).

	Ceiling	Floor	Hab.	MRP	MDP
<i>Arrow</i>	376	161	181	179	183
<i>ATM</i>	733	130	180	182	175
<i>Chimmy</i>	509	157	177	173	177
<i>Elias</i>	697	150	174	170	173
<i>Freddy</i>	733	175	214	224	201
<i>Gareth</i>	468	109	167	166	175
<i>Kel</i>	509	159	213	215	208
<i>Kevin</i>	666	145	177	171	184
<i>Lizard</i>	646	195	207	208	200
<i>Marsh</i> [†]	498	119	174	183	162
<i>Oz</i> [†]	501	96	113	116	110
<i>Panda</i>	467	137	179	179	188
<i>Raine</i>	499	191	194	195	183
<i>Renaud</i>	397	158	198	194	201
<i>Saint</i>	458	226	228	226	230
<i>South</i>	540	164	182	173	177
<i>Ty</i>	394	110	170	161	189
<i>Wizard</i>	600	167	206	206	206

Table 4.3: Starting F0 ceiling, floor, habitual pitch, MRP, and MDP for all speakers in Hz.

is noteworthy, as their initial habitual pitch was, depending on task, either at or extremely close to floor (MFD=0.08), suggesting that their speaking voice pre-HRT was as low as they could comfortably reach. Ty and Gareth had the highest initial MFDs, 7.62ST and 7.32ST respectively. We also see Raine with a lower reported initial floor (191Hz) than MDP (183Hz), a discrepancy seen in others throughout the study and addressed in §4.4.1.

Many participants described vocal techniques they had heard of or attempted in order to alter their voice, with sources ranging from informal voice training groups to online influencers, videos, and blogs. These included physical exercises of the neck and throat, practice to increase awareness of one’s voice such as recitation or humming, and various breath support techniques used by singers. Apart from intonation, voice alteration focused largely on pitch, predominantly involving a conscious effort to remain in the lower part of one’s register.

To quantify these efforts, we cross-referenced the satisfaction survey questions Q16 ("I try to speak in a masculine way"), Q19 ("I try to speak in a male way") and Q21 ("I intentionally try to deepen my voice"). These responses and corresponding interview discussions will be more deeply explored in §6, including a discussion of how speaking in a particularly-

gendered "way" is defined by speakers, though Wizard's reflection on the topic provides a comprehensive summary of the kinds of efforts made (Wizard S1|W1).

I feel that I just, every now and again, check in, and listen to myself talk, and keep it down sort of in my throat. I read one way to sound more masculine [...] people who grow up as dudes, they'll talk with their chest voice, so I tried to be a little more conscious of that. And I know that's kind of along a singer's vocabulary, a little bit. [...] I kept seeing posts of, like: "Girls will always end their sentences in a question" [...] so if you end your sentences with, like, towards a *down* instead of an *up*, you'll sound more masculine. So it's just, um, little stuff like that I try to do. (Wizard, S1|W1)

Responses to these questions were encoded numerically from -2 ("strongly disagree") to 2 ("strongly agree"). A linear regression comparing these responses in the initial session to speakers' initial pitches finds no correlation, either with habitual pitch overall, habitual pitch of individual tasks, or MFD. The reported implementation of voice-deepening techniques pre-HRT (or early in the process of HRT) does not therefore appear to have a measurable correlation with overall pitch. There are two caveats to consider; first, while some participants reported consistent efforts to lower their pitch, others did so only contextually. Several described greater efforts to deepen their voice around strangers, paying less attention to pitch in LGBTQ-friendly spaces or with queer interlocutors (Elias S1|W0).

I definitely try to speak, uh, in a more masculine way, like, at work for example. [...] With people I'm more comfortable with, I don't try as hard to lower my voice, or like, speak with certain cadences. (Elias, S1|W0)

As an LGBTQ stranger, it is uncertain how my presence as interviewer may have influenced intentional pitch change, particularly during the first session, and it is entirely possible that speakers simply were not enacting those reported techniques during this initial interview. Additionally, minimal quantitative effect of efforts to lower pitch should not

be taken as indicative that vocal training is unhelpful in achieving a more masculine voice pre-HRT. Rather, it suggests that such efforts should be focused not solely on pitch, but on other techniques currently suggested by vocal coaches and speech-language pathologists for transmasculine vocal therapy.⁴

4.3 Change in habitual pitch over time

Initial analysis of habitual pitch was done via a mixed-effects linear model to determine overall significance of time on HRT; mean frequency in Hz was the dependent variable, weeks on HRT and task (read passage versus image description) were fixed effects, and speaker was a random effect. Similar models were separately created for pitch floor, ceiling, and range, omitting task as an effect. A final model examined participants' MFD, defined as their habitual pitch (across all tokens for a given session) minus their pitch floor; this included task as a fixed effect, as distances from MDP (image-description mean) and MRP (read-passage mean) were included separately.

4.3.1 *Cross-participant pitch change*

Across all participants, we see a significant effect of weeks on testosterone, for habitual pitch ($\beta=-0.857$, $p \leq 0.001$). Individual speakers universally exhibited the characteristic drop in pitch expected from the effects of HRT. The expected negative correlation was found between weeks on HRT and habitual pitch in all cases (Table 4.5), though some participants exhibited markedly stronger effects than others. No significant difference was found between the read-passage and image-description tasks across all speakers. A task effect was seen, however, in several participant-specific linear models (including only weeks on HRT and task as fixed effects), the results of which are discussed in §4.4.

4. Many such techniques focus heavily on resonance (see §2.2.2 over pitch, working to encourage lowering in the larynx and base of the tongue (Block, 2017; Gelfer and Mordaunt, 2018); these may be predicted to affect the vowel space, see §5.

The smallest effect of time ($\beta=-0.222$) was seen in Oz[†]. The next-smallest effects were, in close proximity, those of Arrow, Chimmy, Gareth, Kevin, Marshmallow[†], Southern Boy, and Ty ($\beta=-0.420-0.518$). The strongest effects ($\beta>1$) were seen, in order, in Saint, Freddy, Lizard, Wizard, Elias, Kel, and Renaud ($\beta=-1.722-1.189$). Speakers saw an average habitual pitch drop of 56.4Hz, amounting to 6.29ST over the course of the study, collated over both tasks (Table 4.6). This is comparable to the 6.4ST decrease observed by Irwig et al. (2017), though variation was high (SD=2.40). The smallest differences between initial and final sessions were seen in Marshmallow[†] (-18Hz, -2ST) and Oz[†] (-21Hz, -3.56ST); the next-smallest by a substantial margin was Arrow (-23Hz, -2.35ST), followed by Chimmy (-38.5Hz, -4.31ST). Saint saw the greatest lowering of habitual pitch (-100Hz, -10.13ST), notable as he had the highest initial habitual pitch of all participants.

We must take caution in reporting change in habitual pitch as statistically significant. While an effect of time on HRT on F0 is present, month-to-month natural variation of habitual pitch averages 2.74ST as determined by Coleman and Markham (1991); this same work suggested a ± 3 ST as the benchmark for change in habitual speaking pitch outside the realm of typical variation. By this metric, two speakers (Arrow and Marshmallow) did not see overall change in habitual speaking pitch that was outside the range expected from month-to-month variation, though each did exceed this range in one of the elicitation tasks (image description and read passage, respectively). This corresponds with Marshmallow's otherwise lower change, but is less expected for Arrow, who completed the full 12 sessions. Across the 16 participants who completed all 12 sessions, mean final-session habitual pitch was 127Hz, ranging from Elias (106Hz) to 157Hz (Arrow). Marshmallow and Panda, with 7 and 10 sessions each, ended with pitches of 155Hz and 128Hz, respectively. As expected, every speaker reached Ziegler et al.'s 185Hz "ambiguous" target pitch, across both tasks.

While we had predicted also that 80% of participants would reach the 131Hz "cisgender male normative" target, only 11/18 (61%) did so, with the remaining 7 consisting of Arrow,

		1	2	3	4	5	6	7	8	9	10	11	12	Change (ST)	
Arrow	MRP	179	138	174	172	166	149	153	142	160	141	146	145	-3.64	-2.35
	MDP	183	186	185	171	183	153	170	157	161	148	162	174	-0.98	
ATM	MRP	182	146	125	133	127	121	123	128	132	131	130	117	-8.33	-7.91
	MDP	175	141	145	127	134	140	117	139	120	127	126	112	-7.65	
Chimmy	MRP	173	166	161	158	164	157	144	149	149	149	143	133	-3.86	-4.31
	MDP	177	171	171	165	170	166	160	170	155	164	160	146	-4.55	
Elias	MRP	170	179	166	165	149	136	139	120	114	126	111	108	-9.37	-8.58
	MDP	173	166	144	174	163	167	123	123	124	118	112	104	-7.85	
Freddy	MRP	224	211	196	204	194	175	153	155	154	143	144	133	-6.07	-7.60
	MDP	201	199	196	212	187	196	162	185	161	146	147	140	-9.03	
Gareth	MRP	166	158	128	123	108	104	110	109	105	102	154	126	-5.68	-5.15
	MDP	175	165	143	115	119	114	119	124	143	117	117	124	-4.77	
Kel	MRP	215	200	198	171	171	155	151	150	145	132	158	150	-7.47	-6.77
	MDP	208	191	190	165	146	137	131	167	156	125	141	141	-6.23	
Kevin	MRP	171	163	160	148	155	137	124	135	134	128	134	137	-5.01	-4.43
	MDP	184	167	143	147	141	136	128	137	146	138	137	133	-3.84	
Lizard	MRP	208	196	190	183	177	162	153	144	136	130	128	121	-9.07	-9.15
	MDP	200	182	176	178	171	165	146	133	139	139	125	127	-9.38	
Marsh [†]	MRP	183	173	149	166	160	146	152						-0.53	-2.00
	MDP	162	162	147	161	166	147	163						-3.21	
Oz [†]	MRP	116	100	98	101	99	98	96	94	96	93	95	95	-3.82	-3.56
	MDP	110	96	96	90	94	99	86	87	92	92	95	89	-3.46	
Panda	MRP	179	175	148	145	140	136	132	136	140	127			-5.77	-5.81
	MDP	188	164	160	154	141	129	126	132	139	112			-5.94	
Raine	MRP	195	182	165	150	145	140	129	116	127	129	130	122	-9.42	-8.76
	MDP	183	173	151	132	139	134	122	104	119	110	119	114	-8.12	
Renaud	MRP	194	192	166	159	148	137	125	134	126	131	135	139	-8.87	-7.28
	MDP	201	177	161	159	143	145	117	124	126	127	137	122	-5.77	
Saint	MRP	226	194	173	178	166	153	151	147	134	142	129	132	-10.90	-10.13
	MDP	230	211	168	178	141	152	144	148	131	135	119	124	-9.31	
South	MRP	173	187	145	159	142	148	163	141	135	140	136	138	-5.34	-4.67
	MDP	177	175	146	161	134	134	167	144	148	138	140	147	-3.91	
Ty	MRP	161	139	143	134	125	129	113	115	127	123	127	119	-7.55	-6.18
	MDP	189	136	143	130	126	141	120	118	130	126	138	119	-5.23	
Wizard	MRP	206	199	206	201	204	180	167	134	135	133	125	132	-9.27	-8.51
	MDP	206	204	195	203	207	186	186	135	130	131	113	121	-7.71	

Table 4.4: MRP and MDP in Hz for all participants across sessions 1-12, final change in ST for each task, and final change in ST overall.

Chimmy, Freddy, Kel, Kevin, Marshmallow[†], and Renaud. Ziegler’s own analysis warns that "59.9% of the total variance ... in failure rates for the [131Hz] target is explainable as differences between the studies themselves (e.g., in terms of protocol, targeted population, or testosterone administration). Stated differently, each of these studies might not be measuring

Speaker	Effect of Weeks (All)	Effect of Weeks (0-13)	Effect of Weeks (0-26)	Effect of Weeks (13-39)	Effect of Weeks (40+)	Effect of RP (All)
<i>Arrow</i>	-0.420**	–	-0.618	-0.692	-0.051	-12.869**
<i>ATM</i>	-0.655**	-5.407**	-2.071**	-0.222	-0.799*	0.591
<i>Chimmy</i>	-0.484**	-1.031	-1.151**	-0.434**	-0.610**	-10.642**
<i>Elias</i>	-1.299**	-0.495	-1.038**	-2.078**	-0.848*	0.399
<i>Freddy</i>	-1.697**	-0.979	-2.003**	-2.202**	-1.997	0.725
<i>Gareth</i>	-0.474**	-4.640**	-2.544**	0.204	1.245*	0.229
<i>Kel</i>	-1.227**	–	-2.294**	-2.392**	-0.215	7.904*
<i>Kevin</i>	-0.458**	-2.374**	-1.432**	-0.877**	-0.087	0.597
<i>Lizard</i>	-1.518**	-2.372*	-1.992**	-2.105**	-0.373	5.629*
<i>Marsh</i> [†]	-0.451*	-3.087**	-0.711	-1.660*	–	0.206
<i>Oz</i> [†]	-0.222**	–	-1.695**	-0.772**	-0.063	6.043**
<i>Panda</i>	-0.857**	-2.772**	-1.843**	-0.596*	-1.427	0.314
<i>Raine</i>	-0.900**	-3.451**	-2.976**	-0.855**	0.292	10.031**
<i>Renaud</i>	-1.189**	-1.719	-2.964**	-2.021**	0.332	0.169
<i>Saint</i>	-1.722**	-6.032**	-3.313**	-1.594**	-1.014*	8.037*
<i>South</i>	-0.482**	-3.941**	-2.447**	0.398	-0.201	0.296
<i>Ty</i>	-0.518**	–	-2.459**	-0.930**	0.135	-3.850*
<i>Wizard</i>	-1.481**	-0.391	-0.547*	-3.061**	-0.375	0.683

Table 4.5: Regression coefficients (β) for the effects of weeks on HRT and read-passage task on habitual speaking pitch. Significant effects are in bold; stars indicate level of significance (* $p \leq 0.05$, ** $p \leq 0.001$). Negative β for MRP effect indicates lower habitual pitch during the read-passage task; positive β indicates lower pitch during image description.

the "same" failure rate in the "same" population" (Ziegler et al., 2018, p39). Thus, we cannot conclude with certainty that this level of deviation from the targets is unexpected, nor that this study's participants are unusual in their relatively high final mean F0s.

Finally, habitual pitches in participants' final sessions are not necessarily the lowest attained throughout the study. The majority of participants (13/18) showed lower means (either MDP, MRP, or both) in sessions other than their last; this may be attributed in part to contributing factors such as time of day, hydration, illness, and in part to the expected month-to-month variation as described by Coleman and Markham (1991).

4.3.2 Individual patterns in pitch change

We can identify for many participants a sharp drop in F0 in the first several months of HRT followed by a levelling-out or slower descent, as expected both from prior work and from the

expectation shared by participants that this is where most vocal change would take place (Damrose, 2009; Deuster et al., 2016; Nygren et al., 2009; Irwig et al., 2017). However, this was not universal, and indeed we see evidence that transmasculine speakers exhibit a variety of distinct patterns in F0 change. A series of mixed-effects linear models were constructed in the same manner as the first, on subsets of the data corresponding (assuming ~ 4.3 weeks per month) to the months in which Irwig observed these patterns: Weeks 0-13 (months 0-3), weeks 0-26 (months 0-6), and weeks 13-39 (months 3-9), as well as an additional weeks 40+ to examine change after the 9-month mark (Table 4.5, columns 3-6).⁵ From these results, we broadly describe several patterns of change found, represented visually in Fig. 4.2.⁶

Sharp descent: Speaker shows the strongest lowering of F0 in habitual pitch between weeks 0-13. This is the pattern most commonly assumed in discussion of HRT's effects on pitch, and the one expected by the majority of participants. It is indeed common, seen in 9/18 speakers (ATM, Gareth, Kevin, Marshmallow, Panda, Raine White, Saint, and Southern Boy). This is the characteristic cliff-and-plateau pattern of change, though whether speakers indeed reached a "plateau" after this point or continued to descend varied. ATM, Gareth, and Southern Boy plateaued, showing no significant change between weeks 13-39; the remainder continued to see pitch descend, but at a slower rate.

In this slow-descent phase, there were numerous instances of habitual pitch measuring higher than a preceding month. Many of these "bounces" back to a higher F0 may be explained as the result of either specific circumstances affecting the voice (e.g. participants recovering from illness) or as within the expected range of month-to-month variation in pitch as described by Coleman and Markham (1991).

5. Note that not all participants had more than one session within the first 13 weeks on HRT.

6. A maximally comprehensive analysis would measure significance in pitch change per pair of months; such an analysis was not performed here as participants varied in distance between sessions to the extent that comparisons between each pair of months (e.g. the second and third, where Deuster et al. (2016) found the greatest significant difference in ST) could not be guaranteed to be available for all speakers, thus the use of longer multiple-month ranges in the manner of Irwig et al. (2017).

Delayed descent: Speaker shows no significant lowering of F0 between weeks 0-13, with the strongest lowering instead seen either more broadly in weeks 0-26 or in weeks 13-39. This includes Chimmy, Elias, Freddy, Renaud, and Wizard; it is the same cliff-and-plateau pattern as above, but one in which the initial change takes place later; similarly, some speakers plateaued (with Freddy, Renaud, and Wizard showing no further significant change in weeks 40+) while others continued with a slower descent (Chimmy, Elias). This pattern in particular has implications on satisfaction; the expectation that voice change will be swiftly noticeable may lead those showing delayed descent to concern that they will not see the expected effects.

Steady descent: Speaker shows F0 lowering evenly distributed through the first 40 weeks. This is the case for Arrow, Lizard, and likely Kel, though Kel had only one session within the first 13 weeks. Lizard showed the highest effect of the first 0-13 weeks ($\beta=-2.372$), but only just; when compared to the 13-39 week effect ($\beta=-2.105$), more gradual change is seen. Arrow, notably, saw no significant change in any of the week divisions, but did see significant F0 lowering overall. These speakers were aware of the gradual nature of this change, describing mixed feelings in their final session (Arrow S12|W71), as while greater change would have been preferred, the slower pace allowed for more comfortable adjustment.

Even though I was like, "Oh, I haven't gotten many changes," I at the same time was like, "Well, at least it's been more gradual, and more within my control." 'Cause that's how I feel about my voice, which [...] hasn't dropped, like, *a lot* a lot. And I'm like, gradual as it is, it's fine. I feel like the way it is, it probably just made it a bit easier.

(Arrow S12|W71)

There is the additional possibility of stepwise descent (as identified by Papp (2012), in which speakers' pitch exhibits multiple separate "cliffs" and "plateaus" in succession. Without closer examination of month-to-month change, this may be indistinguishable from sharp or steady descent, though cursory examination of the habitual pitches of Kel and

Lizard may point to this pattern. Very few speakers (ATM, Chimmy, Elias, Gareth, and Saint) continued to exhibit significant F0 change after 40 weeks, with most having reached a plateau by this point.

Notable here is that Gareth's change during this section was positive, as a result of several sessions in which either their MDP or MRP, though not both, was unusually high (see Table 4.6). This may be reflective of the "curved-J" pattern described by Papp, wherein some speakers showed a pitch increase near or after the end of their first year on HRT, which could not be explained wholly physiologically. Too few speakers showed significant change during this time to be certain, though it is worth noting that several other speakers (Raine White, Renaud, and Ty) did show weak positive trends in the 40+ week period. Overall, these results corroborate those of Irwig et al. (2017) and support the assertion that not all transmasculine individuals on HRT experience significant pitch lowering within the first 3 months. We instead see substantial variation in the path that speakers' voices take on the way to a lower habitual pitch; while many do exhibit the characteristic F0 plummet, this process starts immediately for some and is delayed for others, while still others instead show a more gradual lowering across the entirety of the first year.

4.3.3 *Change in pitch range, ceiling, and floor over time*

Participants as a whole exhibited lowering of both ceiling ($\beta=-2.996$, $p\leq 0.001$) and floor ($\beta=-0.854$, $p\leq 0.001$) over time on HRT (Figure 4.3). These did not lower in parallel for most participants; instead, while floor descended relatively gradually, ceiling tended to drop sharply as participants lost access to their upper register. This seemingly resulted in an overall narrowing of vocal range as measured in Hz ($\beta=-2.156$, $p\leq 0.001$). However, conversion to a semitone scale reveals that this narrowing is not psychoacoustically present; i.e. participants' *perceived* vocal ranges typically maintained their width. We will examine first range (in Hz and ST), then floor, mean-floor distance, and finally ceiling.

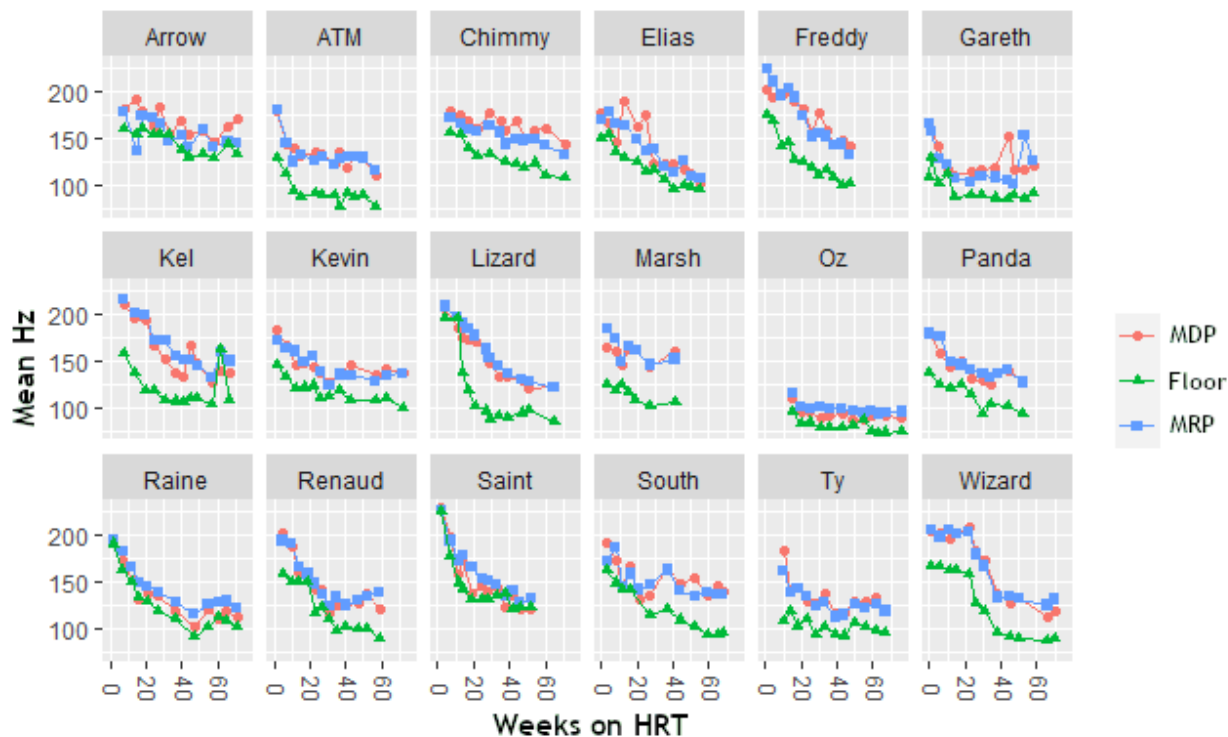


Figure 4.2: Habitual pitch and floor changes over weeks on HRT (in Hz).

Change in range

There is a significant narrowing of vocal range in Hz when measured across all participants⁷ ($\beta=-2.156$, $p\leq 0.001$). Individually, 11/18 speakers exhibited this lowering (ATM, Chimmy, Elias, Gareth, Kel, Lizard, Panda, Raine White, Saint, Ty, and Wizard) while the remaining 7/18 saw no significant change in range; these were individuals who showed either high month-to-month variability or relatively little drop in pitch ceiling (see Fig. 4.3).

However, conversion of ceiling-floor distance in Hz to a semitone scale (ST) and repeating the analysis reveals insignificant overall change in vocal range size on a psychoacoustic level across all participants ($\beta=-0.013$, $p=0.26$, see Table 4.6). These results support the argument that, although pitch range as measured in Hz does narrow over time on HRT for most transmasculine speakers, overall perceived vocal range typically does not.

7. Note that there is no pitch measurement for Chimmy Session 6, or for Panda Session 8.

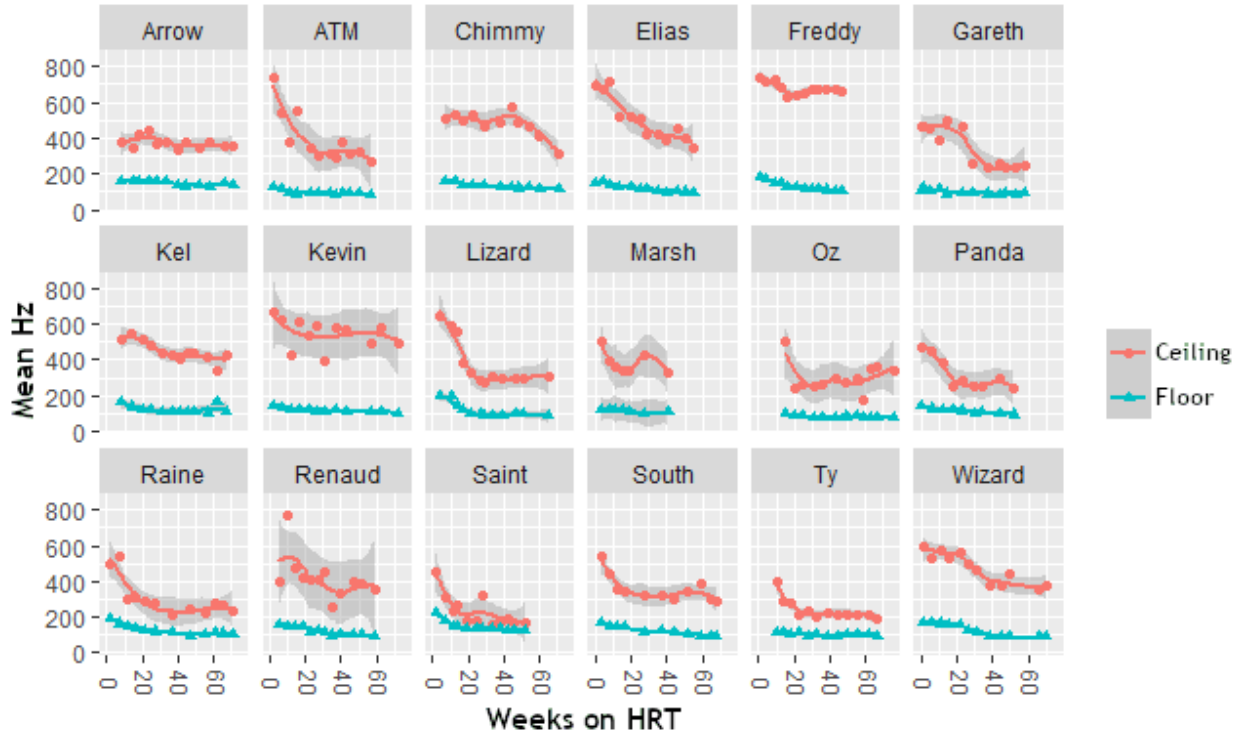


Figure 4.3: Change in pitch range over weeks on HRT (in Hz).

Of the 11 participants who saw vocal range narrowing in Hz, only three (ATM, Gareth, and Ty) exhibited significant range narrowing in semitones. The majority of participants, though losing access to their upper register, did see expansion in the lower register that resulted in an overall maintaining of perceived distance between lowest and highest notes; i.e. those with a three-octave range pre-HRT maintained a three-octave range after the first year of HRT, only shifted downwards. This contradicts the sentiment that HRT results in a universal narrowing, such that those who sang pre-HRT will have to "give up" their range over the course of transition.

That said, three participants did experience a psychoacoustically significant narrowing of the vocal range, with ATM, Gareth, and Ty seeing a reduction of 8.71, 8.14, and 10.75ST respectively. This is a substantial decrease, amounting to approximately 3/4 of an octave (12ST) of range. It is uncertain what the underlying cause of this reduction is for these participants. There is evidence that some part of the measurement of range is affected by

Speaker		Range (Hz)	Range (ST)	Notes	MFD (Hz)	MFD (ST)	Effect of MRP (β)	Floor Change	Ceiling Change
Arrow	First	215.01	14.71	<i>F#4-D#3</i>	20.15	2.05	-13.88	-27.44 Hz	-16.30 Hz
	Last	226.15	17.19	<i>F4-C3</i>	24.63	2.94		-3.24 ST	-0.77 ST
ATM	First	602.40	29.91	<i>F#5-C3</i>	50.03	5.63		-52.78 Hz	-469.36 Hz
	Last	185.82	21.20	<i>C4-D#2</i>	36.22	6.65		-9.01 ST	-17.72 ST
Chimmy	First	352.58	20.39	<i>C5-D#3</i>	19.61	2.04	-11.12	-47.4 Hz	-194.13 Hz
	Last	205.85	18.31	<i>D#4-A2</i>	28.55	4.01		-6.23 ST	-8.30 ST
Elias	First	546.51	26.56	<i>F5-D3</i>	23.24	2.49		-54.56 Hz	-349.31 Hz
	Last	251.76	22.32	<i>F4-F#2</i>	10.58	1.81		-7.81 ST	-12.05 ST
Freddy	First	557.98	24.79	<i>F#5-F3</i>	38.83	2.47		-72.67 Hz	-77.24 Hz
	Last	553.41	32.14	<i>E5-G#2</i>	35.52	5.15		-9.28 ST	-1.93 ST
Gareth	First	358.83	25.18	<i>Bb4-A2</i>	57.55	7.32		-16.95 Hz	-220.99 Hz
	Last	154.79	17.04	<i>B3-F#2</i>	31.26	5.05		-2.92 ST	-11.06 ST
Kel	First	350.15	20.17	<i>B4-D#3</i>	54.11	5.08		-51.44 Hz [†]	-78.48 Hz
	Last	323.11	24.05	<i>A4-A2</i>	36.72	5.10		-6.78 ST [†]	-2.90 ST
Kevin	First	520.87	26.39	<i>E5-D3</i>	31.54	3.41		-45.46 Hz	-171.38 Hz
	Last	394.95	27.75	<i>B4-G2</i>	37.02	5.47		-6.51 ST	-5.15 ST
Lizard	First	450.89	20.75	<i>E5-G3</i>	12.18	1.05		-110.45 Hz	-337.02 Hz
	Last	224.32	22.46	<i>D#4-E2</i>	37.22	6.33		-14.49 ST	-12.78 ST
Marsh [†]	First	372.78	23.95	<i>B4-B2</i>	48.95	5.73		-19.55 Hz	-173.89 Hz
	Last	218.44	19.46	<i>E4-G#2</i>	50.14	6.75		-2.95 ST	-7.44 ST
Oz [†]	First	405.34	28.60	<i>B4-G2</i>	17.09	2.83	4.65	-21.59 Hz	-166.12 Hz
	Last	260.81	26.04	<i>E4-D2</i>	17.65	3.68		-4.41 ST	-6.97 ST
Panda	First	329.47	21.18	<i>Bb4-C#3</i>	42.05	4.62		-45.12 Hz	-228.40 Hz
	Last	146.19	16.44	<i>Bb3-F#2</i>	35.34	5.61		-6.89 ST	-11.63 ST
Raine	First	307.72	16.60	<i>B4-G3</i>	2.90	0.26	10.81	-89.04 Hz	-269.72 Hz
	Last	127.04	13.99	<i>Bb3-G#2</i>	15.04	2.38		-10.85 ST	-13.47 ST
Renaud	First	238.16	15.89	<i>G4-D#3</i>	39.62	3.87		-68.50 Hz	-40.31 Hz
	Last	266.35	23.85	<i>F4-F2</i>	40.38	6.43		-9.81 ST	-1.86 ST
Saint	First	230.99	12.17	<i>Bb4-A3</i>	1.03	0.08		-102.64 Hz	-285.69 Hz
	Last	47.94	5.66	<i>F3-B2</i>	3.44	0.47		-10.44 ST	-16.95 ST
South	First	375.93	20.64	<i>C#5-E3</i>	18.42	1.84		-68.00 Hz	-251.32 Hz
	Last	192.61	19.08	<i>D4-F#2</i>	43.45	6.47		-9.28 ST	-10.85 ST
Ty	First	284.80	22.17	<i>G4-A2</i>	60.60	7.62		-12.43 Hz	-206.44 Hz
	Last	90.79	11.42	<i>F#3-G2</i>	21.89	3.52		-2.08 ST	-12.83 ST
Wizard	First	432.89	22.13	<i>D5-E3</i>	38.64	3.60		-77.72 Hz	-223.60 Hz
	Last	287.01	24.89	<i>F#4-F2</i>	36.79	5.97		-10.83 ST	-8.07 ST

Table 4.6: Participants’ initial and final pitch ranges (Hz, ST, approximate notes on a chromatic scale), initial and final MFD (Hz and ST), presence of a task effect in MFD (β for fixed effect of $task=MRP$ on MFD change in Hz over time, significant values only), and overall difference between initial and final floor and ceiling pitches (Hz and ST). Instances where change over time for a given value was significant are shown in **bold**; positive change **bold and italicized**. Floor change for Kel (marked [†]) was significant only with Session 11 omitted; see §4.3.3.

behavioral factors (see the *floor > mean* discussion in §4.4.1), and so it must be noted that these participants' inability to reach a range of comparable width to pre-HRT could be in part the result of such factors (e.g. discomfort with one's higher register leading to not wanting to "push" too high in the pitch task, despite a physical ability to reach higher notes). I would argue, however, that the cause is in part physiological; participants with reduced ST ranges *did* report "pushing" their voice to its limits, only to find them constrained.

Only five participants (Freddy, Kel, Kevin, Oz, and Wizard) showed a final-session range of at least 24ST. This is notable, as 24ST range is expected for most individuals without professional singing training (Itize, 1994; Papp, 2012); prior work on transmasculine voices has indeed found ranges of ≥ 24 ST in speakers post-HRT (Papp's participants averaged 34.40ST following one year), though some case studies have more comparable results (e.g. Damrose 2009's 13.77ST). This may be the fault of an elicitation technique which did not sufficiently encourage minimum and maximum pitch production, particularly during the first session where self-consciousness regarding the vocal tasks is expected to be high. However, several participants did display expected ranges across various sessions (e.g. ATM S1|W8, 29.91ST; Freddy S12|W47, 32.14ST), suggesting elicitation is not the entire cause.

In the case of Ty and Saint in particular, reported final session range is *remarkably* low, under a single octave. While Ty reached this extreme only following significant reduction alongside ATM and Gareth, Saint showed no significant change in range over time, nonetheless exhibiting a final-session range of only 5.66ST (*F3-B2*), less than half an octave. However, we can be certain that Saint's reported range here is to some extent inaccurate, as this session was an instance of mean-floor crossover (see §4.4.1) in which Saint's habitual pitch during the image description task was in fact *below* their reported floor.

Many participants saw their initial lowering in pitch ceiling associated with feelings of "cracking" or "breaking" as higher pitches were no longer accessible in the modal voice. This was sometimes followed by a re-opening of the upper register as falsetto, but not all

participants were able to comfortably access this register by the end of the study⁸. Thus we may hypothesize that reduced ST range is the result of a modal voice that now seems to "stop at the crack". ATM did report, and successfully produce during one pitch task, falsetto notes (at S9|W41), but did not reach comparable pitches during subsequent months.

Notably, two participants (Freddy and Wizard) experienced a significant *increase* in vocal range in ST. Despite showing an overall decrease of 145.88Hz, Wizard's range in Session 12 was wider than in Session 1 by 2.76ST, further illustrating the necessity of a psychoacoustic scale in this context. Freddy, in contrast, showed no significant decrease of range in Hz, maintaining a high pitch ceiling as floor lowered (see §4.3.3); this resulted in a Session 12 range 7.35ST (nearly a full octave) wider than that of Session 1.

Change in pitch floor and mean-floor distance (MFD)

A significant overall lowering of pitch floor was found across participants ($\beta=-0.837$, $p\leq 0.001$, see Table 4.6, column 8; Table 4.7, column 2). The average distance between initial and final floor was -54.65Hz or -7.43ST; the greatest significant drop in ST was Lizard (-14.49ST), who was the only participant with a final floor over an octave lower than initial, while the smallest were Gareth (-2.92ST), and Marshmallow[†] (-2.95ST). Every participant save for Kel and Ty individually exhibited a significant lowering of pitch floor over time on HRT. Kel specifically showed a rapid initial lowering of floor by week 20 in a pattern similar to other participants, but during Session 11 was unable to reach a comfortable pitch floor during the range tasks, instead producing a supposed floor that was in fact *over* their habitual pitch (see Fig. 4.2). This is presumed to be an outlier; and omission of the Session 11 pitch does reveal a significant pitch lowering comparable to that of other participants.

8. Access to falsetto was self-reported and not considered here in depth; future work quantitatively examining falsetto in transmasculine speakers would be valuable, because lesser vocal fold thickness has been observed in speakers during falsetto during speech in modal voice (Sundberg and Högset, 2001), which may interact with the thickening effects of HRT, because it involves regaining a range to which speakers initially lose access and which they may have either embraced or avoided pre-HRT, and because the inability to access falsetto after some time on HRT may, particularly for singers, be a contributor to low satisfaction.

Ty in contrast exhibited high variability in month-to-month pitch floor, with this change unstable until the final four sessions with a trend towards lower F0 over time that only approached significance ($p=0.06$). Though Ty was 10 weeks on HRT in the first session, and thus any change within those first weeks is not represented here, such change *was* seen in Oz despite a 15-week start; thus it is more likely that their lack of significant change was the result of this variability rather than lack of change, particularly given the change seen in habitual pitch. The relatively low drop in ST found for Ty and Gareth is notable for having contributed to their significantly reduced range (§4.3.3).

There was no significant change across all participants in mean-floor distance (MFD) in Hz, either aggregated or per task. There was a slight but significant overall increase in MFD as measured in semitones ($\beta=0.02$, $p\leq 0.001$), over both tasks. However, examining individual participants, we find that the majority (12/18) showed no change in either semitones or Hz (Table 4.6, columns 5-6). Of the six who did, Chimmy and Southern Boy saw an increase both in Hz and ST, Kel saw a slight decrease across both metrics⁹, and Raine White, Renaud, and Wizard saw an increase only in ST. As with range, we here take semitones to be the more perceptually salient metric for distance between pitches. Thus, between their first and last sessions, those seeing MFD increases included Chimmy (2.03ST, SD=1.08), Raine White (2.12ST SD=0.94), Renaud (2.56ST, SD=1.70), Southern Boy (4.63ST, SD=2.19), and Wizard (2.37ST, SD=1.26), for a mean increase of 2.74ST among those who saw change.

Individual patterns in pitch floor change

Aside from overall change in MFD over time, we may also examine the rate at which floor descended relative to habitual pitch. Just as with habitual pitch, floor lowering falls into several patterns across participants. An array of linear regressions may again be used to

9. Although Kel was the only individual to show such a decrease, this change over time only barely reached significance ($p=0.45$); their final session MFD was nearly identical to their first. Omitting the aforementioned unusual Session 11 mean-floor crossover shows no significant change in MFD for Kel ($\beta=-0.04$, $p=0.10$), suggesting that the apparent decrease was the result of this crossover outlier.

Speaker	Effect of Weeks (All)	Effect of Weeks (0-26)	Effect of Weeks (13-39)	Effect of Weeks (40+)
<i>Arrow</i>	-0.493**	-0.321	-0.182	0.241
<i>ATM</i>	-0.599*	-1.882	-0.339	-0.856
<i>Chimmy</i>	-0.708**	-1.789	-0.627	-0.534
<i>Elias</i>	-1.071**	-1.539*	-1.076	-0.472
<i>Freddy</i>	-1.530**	-2.289**	-0.866*	0.645
<i>Gareth</i>	-0.485*	-1.360	-0.019	0.291
<i>Kel</i>	-0.718*†	-2.471*	-1.238*	0.845
<i>Kevin</i>	-0.470**	-1.216*	-0.304	-0.244
<i>Lizard</i>	-1.642*	-6.206*	-1.750*	-0.633
<i>Marsh</i> †	-0.610*	-0.979	-1.150	–
<i>Oz</i> †	-0.236*	-1.220	-0.750	-0.283
<i>Panda</i>	-0.829**	-0.828*	-1.502	-1.048
<i>Raine</i>	-1.013**	-3.283*	-1.126*	0.539
<i>Renaud</i>	-1.352**	-1.812	-2.548*	-0.871
<i>Saint</i>	-1.435*	-3.984*	-0.120	0.329
<i>South</i>	-1.006*	-1.676	-1.014	-0.625*
<i>Ty</i>	-0.231	-0.231	-0.787	0.002
<i>Wizard</i>	-1.423**	-1.208	-3.066*	-0.097

Table 4.7: Regression coefficients (β) for the effects of weeks on HRT on pitch floor. Significant effects are in bold; stars indicate level of significance (* $p \leq 0.05$, ** $p \leq 0.001$). Floor change for Kel (marked †) omits Session 11 response.

reveal how this lowering is distributed over time, examining each of the week groupings (0-26, 13-39, and 40+) (Table 4.7); the week 0-13 group was here omitted as each participant had only a single pitch floor datapoint per session, resulting in too little data for a range of that size. Indeed, due to the limited datapoints, very few effects in the restricted week ranges of time on pitch floor were significant, compared to habitual pitch.

Among those who do show significant change in particular week ranges, we can identify the expected "cliff-and-plateau" pattern, consisting of a floor F0 drop in the first 0-26 weeks followed by a levelling-out or continued slower descent, for 8/18 speakers (Elias, Freddy, Kel, Kevin, Lizard, Panda, Raine White, and Saint). Many other speakers appear to trend towards this pattern, but results were ultimately insignificant. Due to the collapsing of the first 26 weeks together, we cannot conclusively identify whether Elias, Freddy, and Wizard showed the same delayed descent as in habitual pitch, though examination of month-to-month data does suggest floor and mean lowered in parallel for these speakers (Fig. 4.2).

Similarly, though results were again insignificant, we do also see trends towards gradual change in floor for Arrow parallel to their change in pitch; more surprisingly, we see this also in Ty. As said in §4.3.3, Ty showed no significant effect of HRT on pitch floor; his change in habitual pitch, in contrast, shows a clear cliff-and-plateau pattern, though lack of pre-week 13 data makes it uncertain whether this was immediate or delayed. Though his high session-to-session variability in floor makes this change unclear, there is a trend towards a more gradual shift, such that his habitual pitch lowered prior to a drop in floor.

We do see two other notable cases where habitual pitch and floor are not in parallel: Kel and Lizard, who both showed gradual change in habitual pitch but a sharp cliff-and-plateau pattern for floor. The cause of this difference cannot be determined conclusively, but strongly suggests some behavioral component resulting in "buoyancy" of pitch (Gelfer and Mordaunt, 2018); speakers are capable of speaking at lower pitches, but do not. This is particularly exemplified by Lizard, who went from an average MFD of 0.77ST in Sessions 1-2 to an average of 7.78ST in Sessions 3-9 (with a maximum of 9.46ST in Session 5) before levelling out to 5.41ST in Sessions 10-12. Although this later decrease of MFD caused Lizard to show no overall significant change in MFD over time, it illustrates Lizard's dramatic early drop in floor, while habitual pitch takes nearly the entire year to follow.

Change in pitch ceiling

An overall lowering of pitch ceiling over time on HRT was found across all participants ($\beta = -2.995$, $p \leq 0.001$, see Table 4.6 column 9), as expected. However, this lowering individually was found only in 13/18 participants, with a subset (Arrow, Kevin, Marshmallow, Oz, and Renaud) seeing no significant change in ceiling, either as measured in Hz or in semitones. The average distance between initial and final ceiling was 10.81ST, with the highest significant change from ATM (17.72ST) and the lowest from Kel (2.90ST). In general, pitch ceiling followed the same cliff-and-plateau pattern seen in habitual pitch and floor, with a substantial

drop by the 30-week mark that then leveled out (visualized in Figure 4.2). However, individual patterns are more varied; Chimmy maintained steady access to the upper register until approximately week 50, even as his mean and floor pitches dropped more abruptly.

Among those who show no significant change in ceiling, this was in some cases (e.g. Marshmallow[†], Oz[†]) due in part to an early significant decrease in pitch ceiling that later increased again, either abruptly upon gaining access to one's new range past the vocal break (i.e. being able to comfortably produce falsetto) or more gradually as they grew more comfortable exploring that upper range. Others (e.g. Kevin, Renaud) exhibited substantial month-to-month variation in ceiling, as their voices went through phases of cracking. Arrow and Renaud (as well as Kel and Freddy, though their ceiling drop was significant) produced ceilings in their final session within 3ST of their first, showing remarkably little reduction of the upper register even as their floor lowered.

I wish to highlight the case of Freddy, for whom the maintaining of a relatively high ceiling in combination with floor lowering led to an overall increase in range by the end of the study. Freddy is an avid singer, who described great initial concern regarding the effects of HRT on his vocal range and higher register (see §2.1.3 Freddy S1|W1, repeated below). Central here was the dichotomy between a desire for a lower speaking voice (a desire great enough, in combination with the other physical effects of testosterone, to begin HRT) and a knowledge that one's current singing ability would be forever altered, with no guarantee that one would still be able to sing at all.

It was one of my hesitations to start T, was knowing that I would lose my current singing voice. And there's no guarantee that I would be able to sing in any way that's pleasing after that, or that I would get a range back. [...] That was one of my main hesitations, knowing that I would be giving that up to start T. (Freddy, S1|W1)

Similar sentiments were shared across many of the study's singers ("I would rather not be able to see than not be able to sing." - Renaud, S1|W5), who like Freddy continued to

sing over the course of this first year. While overall vocal range was important for many participants, including those who sang frequently, casually, or not at all, it is noteworthy that among those with little to no lowering of pitch ceiling are several speakers who specifically cited singing as a concern. Just because that higher range *can be accessed* does not imply the same *level of control* over that range as was possible pre-HRT. Returning to Freddy in the final session, he describes that lack of control, slow improvement, and the trouble with pitch-matching also expressed by other singers (Freddy, S12|W47).

I assume it'll take years to build up the same kind of like quality to it that I had before. Cause I spent years singing with my previous voice to get it to that point [...] but I feel like it's getting better, at least it's getting to the point where I'm able to, like, hit the note a little better. Whereas I think a few months ago it was like, I don't even know how to hit the note! Like I can hear it in my head, I'm shooting for it but I'm off completely! (Freddy S12|W47)

4.4 Overview of task effects

While no significant effect is found between the read-passage and image-description tasks across all speakers, a task effect is indeed seen in several participant-specific linear models (4.5). The majority (10/18) showed no effect; 3/18 (Arrow, Chimmy, and Ty) showed overall lower habitual pitches during the read-passage task, as initially predicted based on prior work on read vs. spontaneous speech and the markedness of such speech in trans spaces. 5/18 (Kel, Lizard, Oz, Raine White, and Renaud) showed higher read-passage pitches overall, contrary to predictions.

Even for those showing significant difference in either direction, results are inconsistent, as speakers often varied from session to session in the tasks' relative pitches. Examining habitual pitch via an array of paired Welch's T-tests allows a direct comparison of MRPs and MDPs for each participant during each session. Four participants showed significantly

lower MRPs than MDPs in *at least half* of their sessions: Arrow (8/12), Chimmy (10/12), Ty (6/12), and Gareth (6/12). The first three never showed a significantly higher MRP; Gareth did so only in session 11, potentially explaining the insignificance (Table 4.5) of his own overall MRP effect. Freddy, notably, showed significantly lower MRPs in four sessions, despite no overall task effect.

Only one participant, Raine, showed a higher MRP than MDP in half their sessions (6/12), followed by Kel (4/12). Renaud saw this pattern in only two sessions, and Lizard saw no significant difference in means for *any* session, despite exhibiting an overall effect. Thus, while those for whom MRP is lower appear to exhibit this pattern consistently from month to month, those with lower MDPs tend towards a subtler distinction apparent only broadly over time. The precise implications of these results are uncertain, though they contribute to the growing discussion surrounding the ubiquity of read speech tasks, their benefits and limitations. Authors acknowledge that such a task welcomes an inherent performativity; Zimman describes the task as "an opportunity for [speakers] to perform, in a relatively self-conscious way, a gendered speaking style that projects their affiliations with various types of masculinities" (Zimman, 2012, p212).

Regardless of the direction of the effect, the existence of differences in significance between results for the two pitch tasks supports the proposition that habitual pitch in some trans individuals may appear to drop at different rates depending on the task used in measuring it. A lower habitual pitch in read speech has potential implications on satisfaction, as participants may find their voices in public to be higher-pitched, potentially leading to others not gendering their voices the same way they do. Higher read speech poses the opposite issue, where one might be correctly gendered in public but remain unhappy with one's own voice on private recordings or in non-conversational contexts.

We may question, however, how well the spontaneous speech generated by the image description task reflects the specific kinds of interactions in which participants reported dif-

faculty maintaining a comfortable pitch and overall speech style. Not only was the image description task data highly varied in the kind of speech it elicited (§4.4.2), but multiple participants specifically mentioned certain specific styles that they perceived as highly feminine, or as leading others to assume they were women, even after the study. The most prominent of these was the "customer service voice" (Marshmallow S1|W3).

When I'm at work I have that, like, "phone voice" that happens, that customer service voice that kicks in. It's super hyper-feminine, I can hear it, you know. It doesn't sound like me, but I know it's me. (Marshmallow S1|W3)

Though outside the purview of the current study, the prevalence of a customer service-specific speech style in participants' daily lives was tremendous. Over half of participants were in either a service industry or an administrative one in which conversations, either in-person or over the phone, with customers or clients were common. Speakers had an intuitive awareness of several characteristics of this speech style, along with an interpretation of why these characteristics were present; it was described most often as feminine, approachable, non-intimidating, and non-confrontational, and defined predominantly by higher pitch (Chimmy S1|W7).

Part of it is [...] the customer service phone voice. Where I — I don't know where the expectation comes from, but I feel women are expected to sound — to sound childlike, to sound innocent. There's this high pitched, sweetly toned voice that's expected and [...] if an interaction is going south, the pitch tends to go up. (Chimmy S1|W7)

Aside from the character attributes associated with this vocal style, Kel also viewed "customer service voice" as partly a function of volume, where higher pitches were found to be better-understood by customers in louder settings and so he found himself adopting them in those contexts; this bears consideration over the course of HRT, given reports of issues with volume during the first year (Azul et al. 2017, as well as participants' interview

responses, see Chimmy S11|W60 in §3). Kel further related this vocal style to a theatrical performance; while he does have a performing background, he was not alone in drawing this connection. Ty, meanwhile, described use of a "customer service voice" outside of the context of customer service, using the phrase more broadly to refer to a higher-pitched speech style he used with strangers.

As a marked vocal style, "customer service voice" is likely not well-represented by the image description task; while both are forms of spontaneous speech, depending on the service in question the former may be expected to involve rote recitation or the use of a particular script. Though the exact acoustic characteristics of this style are uncertain, it is clearly associated with a higher perceived habitual pitch, and just as clearly a source of stress for some speakers. As such, future work aiming to understand and improve speaker satisfaction with their voices would do well to focus on this style as a point of concern.

4.4.1 *MFD and mean-floor crossover*

In MFD, we find no significant effect of task between either read-passage or image-description mean and floor. Individually, fewer participants showed such an effect here than in habitual pitch: a total of three (Arrow, Chimmy, Freddy) saw a significantly smaller mean-floor distance for MRP; two (Oz, Raine White) saw a larger MRP distance. This is consistent with Arrow and Chimmy using a lower habitual pitch during the read-passage task, and Raine White a higher one; however, neither Freddy nor Oz showed any significant difference, nor did any other participants who showed a task effect in habitual pitch show one here.

We must also address the appearance of mean-floor *crossover*, the phenomenon found for some participants in some sessions where the measured pitch floor (as produced during the pitch range elicitation task) was *higher* than the measured habitual speaking pitch (produced over the course of the read-passage and image description tasks). This was an uncommon occurrence, but is an impossible result if we assume that range elicitation perfectly measured

participants' floor and ceiling; thus, we conclude that it did not, and that there was some confounding effect of the range elicitation task itself at play.

The individuals who exhibited this crossover effect were Arrow (Sessions 1, 6), Kel (11), Lizard (1), Raine White (4, 10), and Saint (1, 9, 11, 12). Speakers varied per instance in whether floor was above MRP, MDP, or aggregate mean habitual pitch.¹⁰ Of this group, it may be relevant that Arrow, Raine White, and Saint also showed some of the smallest overall ranges in Session 1 (14.71, 16.60, and 12.17ST respectively), with only Renaud (15.89ST) also showing a sub-20ST range. As described in §4.3.3, ranges below 24ST are unexpected overall, and Saint's Session 12 5.66ST is particularly notable.

I suggest that this discrepancy may be due to a combination of self-consciousness and self-awareness of one's range, alongside the explicit request by the interviewer for participants to reach their highest and lowest "comfortable" pitches; as such, participants would frequently stop at a point where they began to feel initial discomfort, though they were encouraged to continue if they did not feel they were straining or harming their voice in doing so. For some, it is possible that a pitch floor perceived as uncomfortable when put on the spot is in fact within comfortable habitual speaking range, and not seen as uncomfortable when produced in casual speech. Participants at multiple points throughout the study, expressed a sense they *could* go lower (or higher), but that it would feel painful, awkward, strained, or "fake". In effect, while some participants may feel they cannot reach what they perceive as extremes of pitch, they in fact naturally produce such pitches in conversation.

4.4.2 *Consideration of image description responses*

We may also examine the elicited data from the image description task in particular through a more qualitative lens. Participants varied broadly in how they engaged with this task,

10. Specifics as follows: Arrow [1:MRP, 9:MRP/MDP/agg.], Kel [11:MRP/MDP/agg.], Lizard [1:MDP/agg.], Raine White [4:MDP, 10:MDP], Saint [1:MRP, 9:MRP/MDP/agg., 11:MDP, 12:MDP]. No significant patterns found.

resulting in noticeable inter-participant differences in speech style. These differences are largely unquantifiable and inherently subjective, relying on my observations as interviewer, but they are discussed here as I believe they provide necessary context to the discussion of participant-specific task effects; namely, I posit that the lack of significant task effects for some speakers, such as Gareth and Wizard, is *more surprising* following their consideration.

When presented with the *Cookie Theft* image and instructed to "describe what is happening", most participants briefly described the literal goings-on of the scene in a consistent style and tone. The answers of ATM, Arrow, Freddy, Lizard, Oz, Panda, Raine White, Saint, Southern Boy, and Ty were typically succinct in this way; those of Chimmy and Renaud tended to be longer, but similarly literally descriptive. These speakers occasionally speculated on the thoughts or lives of the people depicted, particularly in later sessions, but an accurate description of the image was the primary focus; this is what was initially anticipated from the task, and it was the majority result.

Elias, Kel, Kevin, and Marshmallow tended towards longer responses, particularly after having seen *Cookie Theft* multiple times. Their descriptions had a similar structure as the speakers above, but were often more descriptive and casual in tone, containing speculation on the thoughts or lives of the people depicted¹¹. Stylistically, these responses were often distinctly comedic. Being asked to describe the same image every month, with no overt explanation given as to why, led to some expressing (typically mock-)exasperation with the task, improvising stories about the scene¹², or artistically reinterpreting the prompt.

Wizard and Gareth stand out in that a literal description of the image was often not the framework of their responses. Wizard followed long tangents, discussing the image's

11. "This is a, like, fifties- sixties-housewife scenario. Um, she is washing dishes and [...] probably dissociating as the sink overflows, and is ignoring the impending chaos that is the child [...] defying gravity on the stool and stealing cookies, and the other child also not questioning that." -Elias S12|W55

12. "We return to the scene of the crime, where we see a woman in a blue dress [...] [T]he plumbing is not fixed. It is not fixed because her husband, who always did those fixing things, she has murdered him and buried him in the backyard, and we can tell because she's staring out the window at his grave." -Kel S9|W49

artistic features and interpretations, and developed over the course of the study an exaggerated frustration towards its aesthetic inconsistency. This resulted in a distinct, playfully emotionally-charged speech style. Gareth experimented with differently-stylized responses and, therefore, a wide range of speech styles. The task served the secondary purpose of prompting discussion of the aesthetics and gender roles presented in it; speakers varied in how and whether they gendered the human figures in the image, and several used it as a starting point to speak about their own experiences with gendered clothing, implicit cultural expectations about femininity and household tasks, and the like. Though such commentary was more often raised through the interviews, the image for some provided a lens through which to do so. Gareth in particular often approached the task in this way.

Their S10|W47 response incorporated onomatopoeia and sound effects into the scene, followed by comments about having discovered, when first studying gender 20-30 years ago, that boys were more encouraged to make such noises as part of play. Another response describing the picture wholly in terms of geometric lines and figures prompted a discussion of silhouettes, and of feeling comfort with the shape of one's body. Though *Cookie Theft*, personal struggles, successes, and tragedies were acknowledged; in this way, the image description task became a tool for self-expression (Gareth S12|W58).

There's a semblance of pristine, mid-century heteronormative privilege and aesthetic purity, belied by underlying chaos that threatens the facade of perfection. It's an antiquated image, more so for you perhaps than for me, because this is the world I was born into, but for younger generations this scene represents an archaic period usually seen through the lens of film. [...] [M]ore than anything, in spite of these four blotches of acid primary colors, what I see are the sounds. I see the sound of the water spilling, [...] I hear the rumbling in the tummy of the smaller one on the ground, requesting help with the treats on the high shelf of the cabinet. The cookie jar. Isn't that what we all want? A treat? (Gareth S12|W58)

I describe the nature of these responses for two reasons; first, to illustrate that they were not uniform, and so comparing acoustic properties between tasks is not identical across speakers. This variety in results was unexpected, but I consider it nonetheless a fulfillment of the task's purpose. The initial intent was to produce less self-monitored speech compared to the read passage; participants were not told this.¹³ No feedback was given regarding interpretation of the task; participants were not instructed, for instance, to speak as naturally as possible. As a result, some responses appeared *more* methodical, with slower speech or long, contemplative pauses; Gareth's more literary-styled responses exemplify this. While the image descriptions may not be "more naturalistic" or representative of conversational speech for all participants, I argue that they are more representative of *variety* in speech.

However, we see no significant differences between tasks in habitual speaking pitch for either Gareth or Wizard. Arrow, Chimmy, Oz, and Raine White, who did see such differences, provided responses that were not subjectively, recognizably different in style from those of speakers for whom no task effect was found. As such, while some speakers do differ between tasks, the content of their responses does not appear predictive of that difference.

The second reason to highlight these responses is simply that there is neither time nor space in the current work to more than briefly touch upon the breadth of topics, observations, and personal experiences that were discussed in the sociolinguistic interview portions of this study. To delve into those conversations in depth would be an ethnography in its own right; I had not anticipated *Cookie Theft* becoming a source of humor, emotion (real or exaggerated), or poetry, but it would be amiss not to acknowledge the role that this particular task played for some speakers in our interviews.

13. While many had come to that conclusion by the study's end, some form of "So what was up with that [expletive] picture?", followed by laughter, was the most common question during the debriefing portion of the final session.

4.5 In Conclusion

Overall, participants showed the expected lowering in F0 across habitual pitch, ceiling, and floor; individually, all showed lowering of habitual pitch, all but Ty showed lowering of floor, and all but Arrow, Kevin, Marshmallow[†], Oz[†], and Renaud showed lowering of ceiling. All speakers' habitual pitches during the final session were below the "gender ambiguous" 185Hz target pitch used by Ziegler; only 61% were below the "cisgender male normative" 131Hz. Change over time for habitual pitch and floor followed several different patterns corresponding broadly with those identified by Irwig et al. (2017); several participants did not experiencing significant change within the first 13 weeks of HRT, while others showed a gradual descent over the entire first year, rather than the expected "cliff-and-plateau" pattern. No correlation was found between habitual pitch or pitch floor and efforts during or before the first session to lower one's voice or to speak in a male or masculine way.

No significant change was seen in mean-floor distance overall, but five speakers (Chimmy, Raine White, Renaud, Southern Boy, and Wizard) saw an increase in MFD over time in one or both tasks, while Kel saw a decrease. This suggests that while most speakers will not see any change in how far their speaking voice sits from their lowest comfortable pitch, a substantial minority of speakers will speak *further* from this pitch. This is not physiologically explained by the effects of testosterone, and rather is almost certainly a behavioral change; once a comfortable speaking range is achieved, speakers appear to maintain that range even as pitch floor continues to lower.

Significant narrowing was found for range in Hz, but this translated to insignificant change in semitones, which is a more accurate representation of human perception of pitch; only three speakers (ATM, Gareth, and Ty) showed narrowing on this scale, while a further two (Freddy and Wizard) saw an increase in range. Regardless of the productive capabilities of participants' vocal tracts, it is their (and others') perception of their voices which ultimately influences satisfaction. Whether range decreases over time on HRT is, as stated, contentious

in the literature, and this study suggests that it does not do so across all participants; though floor lowers consistently, some speakers do maintain access to an upper register, particularly through the use of falsetto. Through this, such speakers may find themselves with a *wider* range following the first year of HRT; this is exemplified by Wizard, whose 145.88Hz decrease in range in fact amounted to a small but statistically significant 2.75ST *increase*.

While vocal range narrowing is thus far from a universal experience on HRT, it is a possible occurrence. The behavioral and psychological component of changing vocal range where it is impacted cannot be understated, with Gareth ultimately expressing some of the most severe distress and uncertainty regarding his voice over the course of the study; though by the end of the year their comfort with their voice improved, he described particularly during sessions 5-7 a sensation of "claustrophobia" or feeling of restriction (Gareth S5|W14).

There's something about not being able to raise my pitch that really makes me feel claustrophobic [...] and so I'm really going through, like, some real psychological discomfort around my voice. Even though I like my speaking voice, I like how my voice sounds, [...] I'm really enjoying, especially in my readings, allowing my voice to be very soft. And to speak in this whole other modality. But there's actually – there's more, there's more to this discomfort that is even deeper, there's [...] certain songs that I used to love to sing along with. Now I don't even want to listen to them, because when I try to sing along to them my voice doesn't do – I can't hit those notes. And it is kind of freaking me out, a little bit. (Gareth, S5|W14)

Though it is likely impossible to associate this (inherently subjective) sensation of vocal claustrophobia to any single acoustic property, were it a consequence of the reduction of range in Hz, we would expect to see a similar sentiment expressed among many other speakers. Instead, I suggest that a better explanation for this experience is reduction of range in semitones. Although most speakers did not see this reduction, we must acknowledge this sensation of restriction in range described by those who did.

That some saw a widening of range due to little change in ceiling is also notable; the precise reasons behind the differing patterns of ceiling change are unknown, as there are numerous interpersonal and month-to-month differences to take into account regarding ceiling in particular. Whether speakers had recently sung, done vocal warm-ups, or talked throughout the day may have had an effect, as could hydration, illness, and a number of other variables aside from gender and the effects of HRT. Additionally, many participants had been singers pre-HRT and continued to sing throughout the first year, adjusting from soprano or alto roles into tenor or baritone ones while maintaining use of their higher register where possible; others spoke about having been uncomfortable trying to reach higher pitches pre-HRT due to dysphoria, and growing more comfortable with that part of their range over time as it dropped.

Finally, the ranges produced by speakers in the final session were unusually low, typically below 24ST, with two speakers (Ty and Saint) showing ranges of less than an octave. While possible over the course of the first year, as the voice is in the process of change and speech outside certain small ranges may be highly uncomfortable, this is extremely unexpected in the final session, which was after 67 and 52 weeks for those speakers, respectively.

CHAPTER 5

VOWEL FORMANTS, VTL, AND HRT

The following is an analysis of the acoustic properties of speakers' vowels, focusing on both individual vowel formants and on an estimation of vocal tract length (VTL). First, overall formant lowering is predicted behaviorally based on the sociological associations of lower formants with maleness or masculinity, such that individuals attempting to speak in a more masculine style affect lower formants as well as a lower F0 (Canessa-Pollard et al., 2012). Second, existing work with trans speakers suggests this overall lowering does to some extent occur, though not universally (Papp, 2012; Zimman, 2012), and that it may be predicated on biological effects of testosterone as well, including an increase in VTL (Buckley et al., 2022; Cler et al., 2020; Hodges-Simeon et al., 2021).

In this study, analysis of vowel formants makes use of the same tokens as analysis of pitch, including both the read-passage and reading description tasks as described in §4. Overall change in vowel formants was analyzed initially via a mixed-effects linear model where frequency in Hz was the dependent variable. Fixed effects variables were weeks on HRT, the formant itself (F1, F2, or F3), and the task (read passage or image description), while random effects variables were the vowel set and speaker. Following this, analyses of change in specific formants, vowels, and speakers was done by removing that factor as a variable and re-running the model on the relevant subset of data.

Incorporating task as a fixed effect allowed any potential task effect to be examined while also aggregating data between the read passage and image description task, creating a total of 46,350 tokens (15,450 per formant F1-F3). This decreased the likelihood of speakers not producing a particular vowel during any given session, largely a concern in the image description task, where word choice was left to the speakers and number of tokens could vary. The total token count was as follows: DRESS (6465), LOT (5874), KIT (5847), TRAP (4515), PRICE (4455), STRUT (3819), GOAT (3813), GOOSE (3024), FOOT (2955), FLEECE (2808),

FACE (2775). The list of potential words from which these tokens were drawn is provided in §4; all were in stressed positions, and the vast majority were in inter-obstruent CVC contexts, though additional tokens violating this constraint (open syllables, approximant-adjacent, rhotic-adjacent) were allowed where their omission would otherwise result in a participant having no viable tokens for a given vowel during a particular session.

Following analysis of individual formants, estimates of participants' VTL for each session were made using the equation below (Stevens 2000), where n is the formant number, F_n that formant's frequency in Hz, and c the speed of sound in air (34,300 cm/s), producing an approximated VTL in centimeters.

$$VTL = \frac{(2n - 1) \cdot c}{4 \cdot F_n}$$

While the frequencies of all vowel formants in Hz correlate with vocal tract length, estimates of VTL that have examined trans male speakers thus far have focused on higher formants (F3, F4) due to their relative resistance to articulatory differences. Hodges-Simeon and the Buckley et al. case study used an average of F3 and F4 (of /ε ɒ/ and /ʌ/, respectively), while the Cler et al. case study used F4 (of /ɑ i æ eɪ/) exclusively. Our own analysis uses F3 exclusively, as F4 was not included in the current dataset, within the vowel set /ε æ eɪ/ (DRESS, TRAP, FACE). This set avoids the constriction of high vowels and lip protrusion of round vowels, which complicate VTL measurement; due to the expected rounding of some tokens due to the COT-CAUGHT merger, LOT was excluded. Data from both tasks were collated to maximize number of available tokens per speaker; means were calculated first within vowel sets (i.e. all /ε/ tokens together) and then across all vowels. The use of F3 in VTL estimation leads to some conflicting predictions; while we expect the latter to increase based on the aforementioned prior work, other evidence suggests F3 to be highly variable over time on HRT, with a trend towards F3 *raising* seen either for specific participants (Papp 2011) or overall (Zimman 2012).

It must be acknowledged that the vowels used in this analysis are not universally monophthongal; PRICE /aɪ/, FACE /eɪ/, and GOAT /oʊ/ in particular are diphthongs in many North American English varieties. Measurements for each formant were taken from the midpoint of the vowel, and so the starting and ending formant frequencies of these diphthongs were effectively collapsed. Additionally, participants differed in the degree to which they exhibited various vowel mergers, including COT-CAUGHT (which merges the low back vowels /ɔ̃ ɑ̃ ɒ/) and PIN-PEN (merging /ɛ/ and /ɪ/ before nasal consonants). While the former is widespread (though not universal) across the US and Canada, the latter is more confined to regional varieties in the southern United States as well as some forms of African American English (AAE) (Labov et al. 2006). Given the regional spread of these participants, I have chosen to aggregate COT-CAUGHT-affected vowels into the LOT /ɑ/ set, while distinguishing vowels affected by more regional-, demographic-, or context-specific mergers such as PIN-PEN.

Furthermore, while normalization of vowel data is standard across sociophonetic research, I have made the choice here not to do so. Normalization is the process of standardizing across physiological differences between speakers, so as to make sociolinguistic differences more apparent; for example, when comparing dialectal variation in a mixed-gender speaker sample, we may examine the social component of that variation without interference from differing vocal tract lengths or mouth sizes. There are numerous methods, of varying levels of efficacy (Adank et al., 2004; Flynn, 2011). However, crucial to this work is the expectation that participants will experience *physiological* change to their vocal tract; thus, as we wish to be able to observe change due to these effects, such methods are not appropriate here.

To guide the course of discussion, we predict: **(1)** that the formants F1 and F2 will show an overall significant lowering across all vowels and participants; **(2)** that there will more participant-specific change in F3, including some instances of raising, but that any pattern of F3 lowering will correspond to **(3)** a significant increase in estimated vocal tract length; **(4)** that open vowels (TRAP, LOT) will show greater F1 lowering than others.

5.1 Change in formants (F1, F2, F3) over time

The initial mixed-effects model revealed a significant effect of weeks on HRT and frequency in Hz ($\beta=0.196$, $p\leq 0.05$), suggesting that summed across all participants, vowels, and formants, time on HRT does have an overall effect on vowel formant frequency. Initially, this pattern across all formants appears to be positive, which on the surface is unexpected; however, the patterns seen between F1-F2 and F3 differ drastically such that that conflating them does not appear to provide a useful metric. The interactions between formant frequency and time are significantly stronger when each formant is analyzed individually, via performing this same analysis on subsets of the data including only those formants. Figure 5.1 shows change in F1, F2, and F3 across all speakers and all vowels with respect to time.

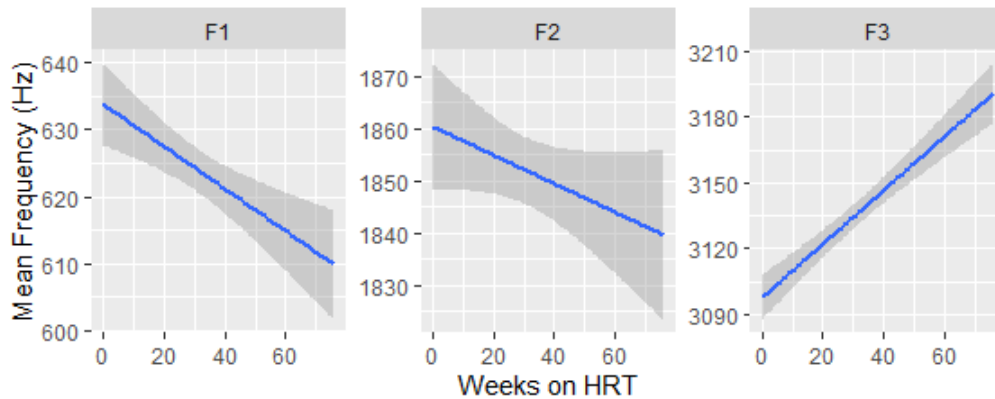


Figure 5.1: Overall formant changes over weeks on HRT.

Time on HRT was found to have a significant negative effect on both F1 ($\beta=-0.50$, $p\leq 0.001$) and F2 ($\beta=-0.35$, $p\leq 0.05$). Both the strength and significance of that effect were stronger for F1. In contrast, a *positive* effect of time was found for F3 ($\beta=1.41$, $p\leq 0.001$), explaining the results of the initial model. The F1 and F2 results are as predicted given prior work with trans speakers' vowel formants, as is the relative strength of the F1 effect. The F3 results are less expected; though an increase in F3 is not unattested (as seen in individual participants within the work of Papp, and across all participants by Zimman), the intensity of this effect is surprising, as it exhibits the highest beta of the three.

5.1.1 *Vowel-specific formant change*

Not all vowels exhibit these effects equally. Collapsed across participants, while significant lowering in F1 and raising in F3 was typical across most vowels, change in F2 was more variable (Table 5.1). Every vowel but FLEECE, FOOT, and GOAT showed an inverse correlation between F1 and time on HRT, and all but FLEECE, FOOT, and FACE showed F3 increasing over time. Most vowels did *not* exhibit significant change in F2, with only FACE, LOT, and FOOT as exceptions; this may explain the comparatively weaker effect of time on F2.

No vowel showed overall raising in either F1 or F2, or lowering in F3; where present, direction of change was consistent throughout. FLEECE uniquely showed no change in any formant¹, while LOT was the only vowel to exhibit change in all three. Overall we see no discernible pattern, in terms of natural classes, in which vowels appear to exhibit formant change over time. Vowel frontness and roundedness do not appear significant. While Papp found a tendency within participants for open vowels to show greater change in F1, this does not appear to be the case here; LOT, though the sole vowel to show three-way change, also exhibits the smallest significant effects across all three formants (Table 5.1).

An argument could be made that the lack of significant F1 effects for FLEECE and FOOT suggests an effect of vowel height (such that, rather than open vowels showing greater change, close vowels may show *less* change), but this is countered by the F1 effect seen in the close vowel GOOSE. Additionally, while no change in GOAT was significant, it did exhibit a trend towards *increased* F1 and F2 over time that is worth noting, as it was the only vowel to show cross-participant raising approaching significance ($p=0.08$ and 0.07 , respectively) in any formant other than F3. The question of patterns in natural classes is revisited in §5.1.2.

1. Data for FLEECE does end prematurely (Figure 5.2), due to relatively few participants having been upwards of 70 weeks on HRT by the end of the study; among those who were, there were insufficient FLEECE tokens for a analysis. This is a consequence both of word choice for one task having been left to participants, and of interference from the use of creaky voice in the other, which interferes with formant analysis. The majority of speakers did provide FLEECE data for all sessions; effect on token count is negligible, and it is unlikely that the lack of change observed in FLEECE can be attributed to the early drop-off of relevant tokens.

Vowel	F1	F2	F3
FLEECE	-0.280	-0.615	0.659
KIT	-0.594**	-0.191	1.246**
FACE	-0.877**	-1.100**	0.493
DRESS	-0.600**	-0.143	1.682**
PRICE	-0.551*	-0.085	2.104**
TRAP	-0.533*	-0.470	1.737**
LOT	-0.363*	-0.902*	1.093*
STRUT	-0.491*	0.148	1.246*
FOOT	-0.387	-1.335*	1.012
GOAT	0.349	1.093	1.671**
GOOSE	-0.653	0.150	1.881*

Table 5.1: Estimated fixed-effects coefficients for effects of weeks on HRT per formant, per vowel, summarized across all participants and both tasks. Significant effects are in bold; stars indicate level of significance (* $p \leq 0.05$, ** $p \leq 0.001$).

These results do contrast with those of Papp, who found overall lowering distributed across all vowels including F3, though specific vowels did show F3 raising in individual speakers (FACE, DRESS, TRAP for one, FOOT, GOOSE for another). Papp’s analysis took into account only two points per participant, zero and one years on HRT. This was not done here as participants varied widely in how long they had been on HRT as of their last sessions, but the differing methods do preclude a direct comparison between the results. Nevertheless, the consistent raising in F3 is surprising, as is the inconsistency with which vowels exhibit the F1-F2 lowering that we do see (Figure 5.2).

5.1.2 Participant-specific formant change

As predicted, participants varied in the extent to which they exhibited F1-F3 change. Most speakers (16/18) showed a significant effect of time on some combination of F1, F2, and F3. The majority showed effects in F1 (12/18) and F3 (10/18) while fewer (8/18) showed an effect in F2 (Table 5.2); in general, participants trended towards higher F3s and some combination of lower F1s and F2s. Only Chimmy, Kevin, and Renaud saw overall change across all three formants. Elias, Ty, and Wizard exhibited no change in either F1 or F2;



Figure 5.2: Vowel-specific formant changes over weeks on HRT, summarized across all participants.

of them, only Elias showed F3 change. All participants with change in F1 saw a negative correlation with time, with the exceptions of Oz and Panda, who showed overall F1 raising and no other formant change. Similarly, all change in F2 was negative except for Kel and Renaud. Where change was seen in F3, it was near-universally an increase in frequency, with the sole exceptions of Elias (whose decrease in F3 was also the only significant vowel formant change seen for that speaker) and Panda.

Given that certain vowels showed greater effects of time on HRT, additional regressions were performed on further subsets of the data examining individual vowels, per formant, per speaker (Table 5.3). We must take the results of this vowel-specific analysis with some care, due to the relatively small number of usable tokens (rarely, as few as 1) produced per session, per participant, per vowel. However, we may make several broad observations: First, not only do vowels not show uniform formant change between participants, but participants who do see formant change do not see it in every vowel (e.g. Panda and Gareth showed

Speaker	Effect of Weeks			Effect of RP Task			
	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>Overall</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>
<i>Arrow</i>	-0.598*	-1.226*	-1.119	-9.899	-24.147*	8.694	-14.262
<i>ATM</i>	-0.564**	-0.634	1.190*	-2.367	42.330*	-31.900	-17.141
<i>Chimmy</i>	-1.088**	-0.895*	1.344**	-15.659	20.497*	-42.422*	-21.771
<i>Elias</i>	-0.222	-0.033	-1.905*	1.825	24.829	29.873	-51.435
<i>Freddy</i>	-2.391**	-1.866*	1.161	-7.603	65.117**	-48.823*	-37.878
<i>Gareth</i>	-0.068	-0.171	2.480**	9.070	-2.914	-15.126	46.436*
<i>Kel</i>	-1.229**	2.792**	0.202	-18.986	28.899**	-148.415**	64.724**
<i>Kevin</i>	-0.385**	-2.892**	1.046*	-7.793	35.615**	-64.409**	6.105
<i>Lizard</i>	-0.413*	-0.594	2.705**	-37.243**	13.883	-70.686**	-52.923
<i>Marsh</i>	-0.014	-2.248*	-0.466	-40.967**	5.748	-87.749**	-40.168*
<i>Oz</i>	1.642*	1.280	-0.132	25.277	45.170	35.320	-5.534
<i>Panda</i>	0.655*	-0.887	-1.231*	-22.329	35.551*	-56.731*	-43.091
<i>Raine</i>	-0.339**	1.093	5.389**	-49.220**	-68.597*	-82.146*	-68.597*
<i>Renaud</i>	-1.521**	1.485*	6.474**	-63.063**	4.733	-106.055**	-85.193**
<i>Saint</i>	-2.925**	0.645	-1.000	-14.715	17.735	-26.733	-30.420
<i>South</i>	0.112	-0.890*	3.070**	-24.236	10.787	-89.728**	7.926*
<i>Ty</i>	-0.398	0.449	-0.660	-11.069	16.949	-52.520*	1.935
<i>Wizard</i>	0.130	-0.455	-0.694	-47.100	-32.415**	-74.030**	-32.742

Table 5.2: Estimated fixed-effects coefficients (β) for effects of weeks on HRT (per speaker, per formant) and of RP task (per speaker, across formants). Significant effects are in bold; stars indicate level of significance (* $p \leq 0.05$, ** $p \leq 0.001$).

no change in F2, Ty none in F2 or F3). Additionally, for a given formant they may see different directions of change depending on vowel (e.g. Lizard’s -LOT and -STRUT, but +KIT and +DRESS for F2); similarly, a given vowel may change in different directions between formants (e.g. Renaud’s F2 -FLEECE and F1 +FLEECE, with no change in -FLEECE for F1). This does not support the theory of HRT affecting the vowel space directly, which as said predicts uniform lowering for all vowels and all formants.

Participants’ overall formant change does not always reflect their vowel-by-vowel patterns. Kel was one of only two speakers to show overall F2 raising, but no individual vowel of theirs exhibited this phenomenon to any significant degree; the same is true for both Oz and Panda in F1. Similarly, the degree of change seen in individual vowels across all participants does not necessarily correspond to the number of participants who saw change in that vowel, for a given formant; here we do see some evidence of certain vowels exhibiting greater change. We return here to the question of natural classes. First, the mid and closed front vowels appear

to show the greatest lowering in F1 (7/18 speakers for FLEECE, 8/18 KIT, 8/18 FACE, 9/18 DRESS). The seven with FLEECE F1 lowering notably included three (Marshmallow, Chimmy, and Oz) who saw no other F1 change. However, as described in Table 5.1, there was no significant change in FLEECE when summarized across all speakers.

Secondly, aside from this group, the vowel with the most frequent lowering seen in F1 is LOT (7 speakers). Here, we may have a corroboration of Papp's findings, where lower vowels appeared to show greater F1 change. Again, however, despite it being common across speakers, F1 change in LOT was still the weakest across all vowels that exhibited it (Table 5.1). Finally, the trend towards a formant increase for GOAT described in §5.1.1 is seen more clearly here; outside of F3 it is the vowel most likely to show raising, with four participants seeing +GOAT for either F1 or F2.²

The last observation is that every participant saw some significant effect of time for in some vowel, in some formant, even if they did not show significant change for that formant overall. However, this should not be taken as evidence of HRT-motivated formant change for all speakers. Even assuming high validity of this per-speaker/per-formant/per-vowel analysis³, I argue these results are more likely to be the result of other natural change over time. There are any number of reasons for Ty's FOOT vowel, for instance, to lower in F1 over the year of the study; fewer for it to do so to the exclusion of all other vowels, and no evidence that change in the vowel space from HRT would result in the latter.

Ultimately, while there is the potential for a more rigorous examination of vowel height and backness to reveal underlying patterns in these changes across participants, we can make no conclusive claims at this time regarding particular natural classes of vowels displaying different formant changes over time on HRT. Lowering does occur in many vowels for most

2. Both GOAT and GOOSE also show relatively infrequent lowering, potentially suggesting some effect of high back vowels specifically, but the current data is not conclusive.

3. Given the high number of comparisons done, and the aforementioned issue of token numbers, we cannot discard the likelihood of false positives.

	Vowels Exhibiting Significant Formant Change over Time on HRT		
	F1	F2	F3
<i>Arrow</i>	-FLEECE, -KIT, -FACE, -FOOT	-FLEECE, -FACE, -DRESS, -TRAP	-FLEECE, -KIT, -FACE, -FOOT
<i>ATM</i>	-FLEECE, -KIT, -FOOT, -GOOSE, +GOAT	-FLEECE, -DRESS	-FLEECE
<i>Chimmy</i>	-FLEECE	-KIT	+TRAP, +LOT, +GOAT
<i>Elias</i>	-KIT, -FACE, -LOT, -STRUT	-FLEECE, -KIT, -FACE	-FLEECE, -FACE, -DRESS, +GOOSE
<i>Freddy</i>	-KIT, -FACE, -DRESS, -PRICE, -STRUT	-DRESS, -FOOT	-DRESS, +PRICE, +LOT
<i>Gareth</i>	-DRESS, -LOT		+FOOT, +GOAT, +GOOSE
<i>Kel</i>	-KIT, -DRESS, -PRICE, -TRAP, -LOT, -STRUT	-KIT, -DRESS, -TRAP, -GOAT	-FACE, -STRUT
<i>Kevin</i>	-KIT, -FACE, -DRESS, +TRAP	-KIT, -FACE, -DRESS, -PRICE, -TRAP, -LOT -FOOT, -GOAT, -GOOSE	+LOT, +STRUT
<i>Lizard</i>	-FACE, -DRESS, -LOT	-LOT, -STRUT, +KIT, +DRESS	+DRESS, +TRAP
<i>Marsh</i>	-FLEECE	-LOT	-KIT, -DRESS
<i>Oz</i>	-FLEECE	-GOAT	-GOAT, +KIT
<i>Panda</i>	-FACE, -DRESS, -TRAP, -STRUT		-PRICE
<i>Raine</i>	-FLEECE, -DRESS, -LOT, -FOOT, -GOAT	-GOAT, +KIT, +DRESS	+DRESS, +PRICE, +TRAP, +LOT, +STRUT, +GOAT, +GOOSE
<i>Renaud</i>	-KIT, -FACE, -TRAP, -LOT	-FLEECE, +GOAT, +GOOSE	+FLEECE, +KIT, +FACE, +DRESS, +PRICE, +STRUT, +FOOT, +GOAT, +GOOSE
<i>Saint</i>	-DRESS, -LOT, -STRUT, -FOOT	-TRAP, -FOOT, +GOAT	-TRAP
<i>South</i>	-KIT, -FACE, -DRESS, -FOOT, -GOOSE, +PRICE, +LOT	-DRESS, -STRUT, -GOOSE, +FLEECE	+KIT, +DRESS, +PRICE, +TRAP, +LOT
<i>Ty</i>	-FOOT		
<i>Wizard</i>	-FLEECE, +GOAT	-KIT	-GOOSE

Table 5.3: Vowel sets showing significant ($\alpha=0.05$) change over week on HRT, per formant, per participant. Direction of change labeled (+) (-).

speakers across F1 and F2 (Table 5.3), but this correlation is not shared across either speakers or vowels. While these results are consistent with prior work indicating high inter-speaker variation in the presence and direction of vowel formant change, they are inconsistent with the prediction of cross-participant formant lowering, as we would expect from a uniform lengthening effect of HRT on the vocal tract. I instead suggest these effects to be the result of speaker-level articulatory change, informed in part by the effects of HRT on pitch.

This has potential implications on satisfaction, as with evidence suggesting that lower overall vowel formants are associated with voices more frequently perceived as male or masculine, we may expect individuals who showed relatively little formant lowering to also show reduced perceived vocal masculinity; testing this is a logical next direction for future work. Vowel change was substantially less salient to speakers than change in pitch, and so there was little qualitative discussion in interviews about the vowel space and speakers' experiences.

5.1.3 *Task-specific formant change*

The initial LMM showed an overall effect of task ($\beta=-0.069$, $p \leq 0.001$), such that read passage formant frequencies tended to be lower. This supports the hypothesis that a read passage task would tend to elicit more "masculinized" speech, being both a situation in which speakers are aware that their speech will be used for acoustic analysis and a task already associated with voice training within the trans community. However, this effect was not seen across all participants, nor was it consistent across formants.

Only five speakers (Lizard, Marshmallow, Raine White, Renaud, and Wizard) showed significant difference in overall formant change between tasks (Table 5.2, column 5). Each saw a negative effect of the read-passage task on F1-F3. Examining individual formants (Table 5.2, columns 6-8), we find that many more speakers saw differences in F1 (9/18) or F3 (6/18), but that direction of these differences varied; lower read-passage formants were not consistent. 12/18 speakers, however, saw a lower read-passage F2; this effect is by far the strongest, and appears the primary contributor to the significant overall effects seen by Lizard and Ty (who showed no F1/F3 differences). This is unexpected; while in part consistent with the idea that lower formants are associated with a more male or masculine voice, such that those desiring such a voice would produce lower formants in the process, we would expect greater uniformity in the formants affected. Instead, any behavioral change in articulation appears to be primarily, but strongly, affecting F2.

The patterns of task effect seen here overlap with but do *not* perfectly match those seen in F0 (§4.4). Arrow, Chimmy, and Ty showed lower read-passage habitual pitches; only Kel, Lizard, Oz, Raine White, and Renaud showed higher ones. Lizard, Raine White, and Renaud both showed an overall RP task effect on formants. Kel and Oz did not; Oz showed no effect on any formant, and while Kel did show individual effects for all three, the negative effect on F2 appears to have cancelled out the positive effects of F1 and F3 when viewing the formants holistically. Arrow, Chimmy, and Ty further differ in their directions of effect for vowels, showing negative, positive, and no effect of the RP task respectively.

I can offer no satisfactory explanation for this phenomenon, which deserves future consideration; this pattern of vowel backing may be a general articulatory characteristic of these speakers' read speech (though if so, it is unattested in literature on this task), or it may indeed be related to how the markedness of read-speech is predicted to affect articulatory performance of gender in the voice. Regardless, the presence of this difference itself calls into question, as does pitch, how reflective read speech is of trans speakers' everyday voices.

5.2 Change in estimated VTL over time

For each session and each participant, an estimation of VTL was made using F3 measurements of the DRESS, TRAP and FACE vowels. A simple linear regression performed on the resulting data shows an overall negative correlation ($r=-0.165$, $p<0.05$) between VTL and time on HRT; this contradicts the prediction that VTL will increase over time on HRT, but is consistent with the F3 raising observed.

A mixed-effects model incorporating speaker as a random variable similarly suggests an extremely slight effect of weeks ($\beta=-0.005$, $p=\leq 0.001$), but inter-speaker variation accounts for the majority of variance, and a set of speaker-specific linear regressions showed most speakers exhibiting no significant change in estimated VTL (Table 5.4). The lowering effect may be attributed to four speakers: ATM ($r=-0.576$), Gareth ($r=-0.894$), Raine White ($r=$

Speaker	F. VTL (cm)	L. VTL (cm)	r
ATM	13.40	14.39	-0.58*
Arrow	14.11	13.81	0.29
Chimmy	14.20	13.58	-0.38
Elias	12.99	13.57	0.36
Freddy	13.56	13.24	-0.46
Gareth	14.42	13.85	-0.89**
Kel	13.79	13.44	-0.11
Kevin	13.73	13.41	-0.40
Lizard	13.19	12.62	-0.36
Marsh	13.90	13.88	0.21
Oz	12.33	12.44	-0.11
Panda	14.05	14.18	0.48
Raine	13.37	12.19	-0.65*
Renaud	13.95	12.77	-0.84**
Saint	13.13	14.34	0.14
South	15.11	13.48	-0.40
Ty	14.62	14.76	0.12
Wizard	13.70	13.52	0.19

Table 5.4: Initial- and final-session estimated VTL by participant, and Pearson correlation coefficient (r) for VTL and weeks on T. Correlation in bold indicates significance (* $p \leq 0.05$, ** $p \leq 0.001$).

-0.647), and Renaud ($r=-0.845$), all of whom showed a significant decrease (0.3-1.2cm) in estimated VTL over time. That these speakers show this pattern is unsurprising, as all four also exhibited F3 raising, though the other four with this raising (Chimmy, Kevin, Lizard, and Southern Boy) did not also show a lowered estimated VTL.

First- and last-session estimated VTLs had a mean of 13.75cm and 13.53cm. The individual lowest and highest estimates in the first session were those of Oz and Ty (12.33-14.76cm), while those of the final session were Raine White and Southern Boy (12.19-15.11cm). These results fall within or below the range estimated for cisgender women, with no participant exceeding this range as some trans and cisgender men of Hodges-Simeon et al. (2021) did. While these results would appear to contradict those of Hodges-Simeon, Cler, and Buckley, several extenuating factors and possible explanations are discussed in §5.3.

As discussed in §2.2.4, there is some evidence that certain kinds of vocal training and voice manipulation techniques, such as laryngeal reposturing, may affect both perceived masculinity of the voice and estimated VTL (Dahl et al., 2022). Though the exact voice

manipulation techniques used, if any, varied between speakers, few reported manual manipulation of the throat and vocal tract in a manner consistent with laryngeal massage or reposturing; furthermore, Dahl et al.'s work did not instruct speakers to lower their pitch during the manipulation studied. However, we might still predict reported intentional manipulation of the vocal tract to have some correlation with estimated VTL.

As with habitual pitch and mean-floor distance in §4, linear regressions were performed comparing estimated VTLs, across participants, to responses to the satisfaction survey questions Q16 ("I try to speak in a masculine way"), Q19 ("I try to speak in a male way") and Q21 ("I intentionally try to deepen my voice") for that section. No significant differences were found over the full course of the study, either across all participants or for the group (ATM, Gareth, Renaud. and Raine White) who saw VTL change. Given our specific focus on initial and final sessions in discussion of satisfaction, these were examined individually. No correlation seen in Session 1, and so reported efforts to increase maleness of masculinity of speech appear to have no effect on VTL pre-HRT.

There was a significant positive correlation ($r=0.48$, $p\leq 0.05$) only between estimated VTL and efforts to "speak in a masculine way" (Q16) during participants' final sessions. However, examining the penultimate and antepenultimate sessions finds no such correlation, nor was one found for either Q19 or Q21, leading to the suspicion that this a false positive⁴. In general, estimated VTL following the first year of HRT also appears not to correlate with efforts to alter the voice. This is notable, as it suggests that those participants who do exhibit significant VTL change do so because of either some physiological effect of HRT not experienced by the majority of speakers, or because of articulatory change not intentionally motivated by efforts to masculinize or deepen one's voice.

4. As participants varied in weeks between sessions, some speakers' final sessions were same number of weeks on HRT as others' penultimate sessions (and so on); as such, while a session-by-session inquiry is useful to gauge overall satisfaction at the end of the study, statistical analyses of particular sessions over multiple speakers are not truly comparing those speakers at the same point on HRT, and must be taken with caution.

5.3 Discussion of vowel formant and VTL effects

Speakers tend to exhibit lowered formant frequencies in one or both of F1 and F2, and a stronger effect of raising in F3; however, this is neither uniform across vowels nor universal across participants. Estimations of vocal tract length using F3 showed no change for the majority of speakers, and a decrease in estimated VTL for four participants. Findings are summarized in Table 5.5. The F3 trend, lack of VTL increase, and degree of inter-vowel and inter-speaker variation do not support the prediction that testosterone HRT results in overall lower formants due to physiological effects on the vocal tract, instead suggesting stronger influence on the vowel spaces of transmasculine voices from behavioral factors (social and physical comfort-motivated articulatory changes) during the first year of HRT.

<i>Ar.</i>	<i>AT.</i>	<i>Ch.</i>	<i>El.</i>	<i>Fr.</i>	<i>Ga.</i>	<i>Ke.</i>	<i>Kv.</i>	<i>Li.</i>	<i>Ma.</i>	<i>Oz.</i>	<i>Pa.</i>	<i>Ra.</i>	<i>Re.</i>	<i>Sa.</i>	<i>So.</i>	<i>Ty.</i>	<i>Wi.</i>	Summary
										x	x							F1 Increase
						x							x					F2 Overall Increase
	x	x			x		x	x				x	x		x			F3 Overall Increase
x	x	x		x		x	x	x				x	x	x				F1 Overall Decrease
x		x		x			x		x							x		F2 Overall Decrease
			x								x							F3 Overall Decrease
x					x							x	x					VTL Decrease

Table 5.5: Summary of quantitative patterns in vowel formant and VTL change.

Existing work on F1-F3 change in transmasculine people has been largely inconclusive due to the disparity in results, with speakers showing at times radically different patterns in which vowels change and to what extent. This has been argued to be indicative that observed formant change is the result of behavioral, sociolinguistically-motivated articulatory shift (Zimman 2012), rather than physiological effects of HRT. I argue in favor of this conclusion, though I suggest also that part of this shift may not be strictly sociolinguistic in origin, but rather the indirect result of physiological change affecting pitch.

If HRT did lengthen the vocal tract in a way comparable to what is seen in cisgender men during typical puberty, we would instead expect uniform lowering of all vowels across all formants Lammert and Narayanan (2015); Vorperian et al. (2019), though the extent of

this may vary between speakers as F0 does; instead, while the trend towards lower F1s, F2s, or both is consistent with expectations given the results of prior work, not all vowels showed lowering, nor did speakers show consistent lowering in the same vowels. Some participants showed significant F1 or F2 raising, either overall (Oz, Kel, Panda, Renaud), or for specific formants and vowels (ATM, Kevin, Lizard, Raine White, Renaud, Saint, Southern Boy, and Wizard). There is weak evidence for some pattern of natural class, as more speakers exhibited lowering in F1 for mid and high front vowels (as well as for LOT, possibly as observed by Papp), but this is inconclusive and must be the subject of more rigorous later analysis.

Of particular note is the near-universal trend towards a raised F3. Prior work on the vowel spaces of trans voices has found some individuals showing F3 raising (Zimman 2012 in particular seeing this pattern when summarizing over all subjects), but its ubiquity here is unexplained. One potential confounding factor to consider is the nature of remote linguistic data collection; as described in §3.1, there are some mechanical concerns with conducting phonetics research across multiple devices, including inconsistency in higher formants, with F3 showing high variability between the devices used by study participants to record. However, this F3 effect was one of variability between different devices, not within individual sessions. While speakers recorded on a wide array of hardware (typically desktop or laptop computers, with built-in and detached microphones of varying quality), this hardware typically remained the same throughout the study; participants rarely switched devices. Were this the source of the effect, we would expect F3 to remain relatively constant over time for each speaker, regardless of interpersonal differences.

Overall, there does appear to be some effect of time on HRT on vowel formants, but it is inconsistent and thus likely motivated not by the direct physiological effects of HRT. The individual changes seen in participants may be socially motivated, such that manipulation of the vowel space is used, consciously or unconsciously, to affect certain gendered styles of vocal presentation. It may also be indirectly physiologically motivated, wherein other

changes to the voice (e.g. F0 lowering) are affecting speech in a way that prompts alteration of the vowel space. Both explanations may account for the different patterns of change seen, and I believe it is likely that both play some role.

If socially informed, this downwards formant trend is likely not a directly intentional choice; speakers tend not to be aware of gendered characteristics in the voice aside from pitch, intonation and to some extent "resonance", though the definition of latter is nebulous. Furthermore, there appears to be no correlation between efforts to adopt male or masculine speech characteristics and this formant change pattern. It may still be the indirect result of such efforts, given the salience of gender in vowel production. Even if not as aware of these differences as they are of pitch, trans speakers may manipulate their vowel space in order to achieve what they perceive as a particularly-gendered vocal style. There is no guarantee that these manipulations would lead to uniform changes in formants, particularly if different formant information is weighted differently in perception of gender in the voice. Earlier work on trans speaker' vowel formants have indicated the same, suggesting that F1 may be "more salient than changes in F2 when it comes to certain kinds of (gendered?) [*sic.*] sociolinguistic information" or presenting "the most likely possibility ... that the true pattern of change is more complex than simply lowering F1 versus F2 versus F3" (Zimman 2012;145).

We have some recent evidence to support this from Leung et al. (2021), who found that listener perception of speaker gender was predicted only by F2 (and F0), while F1 and F3 were instead inversely correlated with perceived *masculinity* of the speaker. This would suggest that speakers showing decreased F1s may experience greater perceived masculinity in their voice, and correspondingly greater "maleness" for decreased F2, and some indirect awareness of this by speakers could be a potential behavioral motivator for the formant changes seen. This still, however, does not explain the consistent increase in F3 (which we would also predict to correlate with decreased perceived masculinity) nor the speakers who saw increases in either F1 or F2. This research is still preliminary, and future perceptual

work on vowel formant and gender as they relate specifically to trans speakers is necessary, but it does appear that individual formants may correspond to perception of gender in the voice, just as vocal tract length as a whole corresponds to gendered speech production.

The discussion is complicated by the fact that vowel formants are highly susceptible to influence from other sociolinguistic factors, such as regional dialect.⁵ Furthermore, given that no participant saw universal formant change across their vowels, this would appear to suggest that not only particular formants but certain vowels carry gendered information more strongly than others. Such findings would not be entirely unprecedented; there is evidence, for example, that pre-puberty differences in vowel formants between male and female (presumably cisgender) children arise incrementally, with gender differentiation in F1 occurring first in back vowels, and F2 first in high vowels (Vorperian et al., 2019).

Alternatively, the vowel changes seen may be unrelated to sociolinguistic expressions of gender in the voice, but rather indirect results of physiological change on HRT. As pitch lowered, participants frequently discussed feeling that they were speaking from a different place in their mouth, holding their jaw or tongue differently. As such, it is possible these changes in vowel space were not the result of conscious attempts to affect socially-gendered cues in speaking, but an articulatory response to a change in comfort. Different results across speakers might thus be explained by the differing extents to which individuals felt the physical need to alter their speech patterns in response to a lowered F0.

Many also described, as F0 lowered, an increased experience of the use of "chest voice". While the acoustic properties of this phonation style are ill-defined, having subtly different definitions depending on vocal pedagogy tradition, there remains some evidence that F1 is associated with this experience of "resonance"; the resonance characteristics of the head and chest cavities respond more strongly to different vowels, with the latter in particular showing greater resonance in open vowels (Cai, 2019) (§2.2.2).

5. A speaker exhibiting the Northern Cities Vowel shift (Labov et al., 2005), for example, may have a lower F1 for TRAP than other speakers; this lower formant frequency is not, however, associated with masculinity.

I strongly suggest that these reports of a shift to more "chestiness" or "chest voice" or "vibration in the chest", however it is phrased by speakers, is indicative in part of some articulatory change prompted by lowered F0. HRT's effects do not preclude speakers from speaking *outside* of "chest voice"; several expressed a newfound ability to switch between a speech style that was lower in F0 and described as "chest" to another that was higher and had markedly different timbre, described variously as "nasally" or "throaty" or "heady" (or, correspondingly, as "speaking in the nose/throat/head").⁶ Though the differences between these styles are not yet acoustically defined, that speakers are largely able to shift between them suggests that this is not a direct result of HRT, but a behavioral change in articulation as a response to the change in pitch; as F0 lowers, formants tend to follow.

Related to this shift in articulation is the sensation of physical discomfort, which is not uncommon as pitch lowers as attested both in prior work with trans speakers (Azul, 2015b) and in the current study. There is evidence suggesting that the overall vocal health of speakers on testosterone HRT is often sub-optimal (Azul et al., 2017). While specific issues of voice quality, stability, and volume are not discussed in the current work, for several months during the middle of the study the changing voice for multiple participants was described as "hoarse" or "strained"; by the end of the year, such negative descriptors were less common, but speakers continued to comment that the changes would require acclimatization. Change in articulation over time may thus be motivated in part by a desire to avoid discomfort.

Aside from individual formants, the patterns seen in estimated VTL also do not conform to the prediction that HRT causes lengthening of the vocal tract in transmasculine speakers. Instead, ATM, Gareth, Raine White, and Renaud exhibited a significant *shortening* of estimated VTL. It is not physiologically possible, given current evidence, for testosterone to cause actual shortening of the vocal tract; it is possible, however, for speakers to affect

6. I further hypothesize that it is this very style which best represents the phenomenon referred to as the "transmasculine accent" (§2.2.3), though this cannot be determined conclusively without a dedicated perceptual study.

vocal tract length (and formant frequencies more generally) through changes in articulation. There are several possible explanations for the VTL results seen here. First, it is possible that the vocal tract does *not* lengthen as a result of testosterone HRT, and that prior formant-based VTL estimations of transmasculine speakers (e.g. Cler et al. 2020; Buckley et al. 2022; Hodges-Simeon et al. 2021) in fact captured formant lowering as a result of behavioral changes in articulation; we might then ask why the current study’s participants did not show these behavioral changes.

It is also possible that formants overall do not provide an accurate method of capturing gender-related VTL differences in transmasculine speakers specifically. The use of F3⁷ in VTL estimation is well-established and has been shown to capture sex differences in VTL separately from speaker height (Turner et al., 2009; Lammert and Narayanan, 2015). Higher formants show reduced sensitivity to articulatory differences, but given the evidence that active work to manipulate the larynx can lead to lower estimated VTLs for transmasculine speakers (Dahl et al., 2022), it is not an infallible means of measurement.

Alternatively, it is possible that VTL *does* increase, but that behavioral change in the vowel space has a greater effect on formants than physiological effects. This suggests that whatever articulatory change causes consistently raised F3 is stronger than any vocal tract lengthening (which then may be partly responsible for the patterns of lowering witnessed in F1 and F2). Finally, it may be that VTL does increase, but participants in this study have not been on HRT long enough to show the expected lowering in F3. Participants in Hodges-Simeon et al. (2021)’s work had been on HRT for a mean of 41.50 months, compared to 14.6 for the current study, and time on HRT was found in that study to be a significant predictor of VTL. As such, VTL may occur, but take years on HRT to become apparent.

I argue in favor of this explanation, as both alternatives require that the differing results between the current study and Hodges-Simeon et al. be attributed to individual behavioral

7. Typically alongside F4, which may be considered in future work.

differences in their respective participant pools. However, I advise caution in the future use of formant-based VTL estimates with trans speakers. Regardless of whether we can expect a VTL increase for speakers in the coming years, the significance of current results over the initial year remain, and they would appear to suggest a VTL reduction which is not physiologically possible. As such, we must question the accuracy of these estimates in the context of a population whose speakers exhibit significant articulatory change in formant frequencies, and for whom there is such substantial variation.

Further work is necessary to conclusively determine the extent to which these changes in vowel formants and VTL are the result of HRT's direct physiological effects versus behavioral change in articulation, and to what degree the latter is informed by physical comfort versus sociolinguistic understanding of gender in the voice. Regardless of the presence or strength of the physiological component, we must acknowledge that some behavioral component does appear to exist; it is this component on which I place the greatest emphasis, not because I believe it to be solely predictive of the vowel formant change seen, but because it is the primary factor over which speakers have some measure of control.

CHAPTER 6

VOCAL SATISFACTION AND HRT

In the following chapter, I discuss the structure and results of the longitudinal study's satisfaction questionnaire. This will focus on the change in responses over time on HRT while taking into account the identities described in the terminology questionnaire, speakers' personal context, and experience as discussed in the interviews.

6.1 Satisfaction metrics and expectations

As described in §2.3.1, I divide statements aimed at gauging speaker satisfaction into four broad categories, based on how they relate to gender and the voice. **General satisfaction** consists of questions which do not explicitly relate to the perception of one's voice as gendered in a particular way; e.g. whether they like their voice, or believe they sound like themselves. **Internal** and **external perception** ask how speakers would personally gender their own voice and how they believe others do so, respectively; the former also includes current efforts to actively speak in a particularly-gendered way. **Desire for change** asks how speakers *wish* their voices sounded, including specific gendered descriptions ("I wish my voice sounded more male") or the desire to undergo vocal therapy.

A list of 43 statements were composed based on these questions; many were inspired by those in prior studies or other questionnaires, including the TVQ, while others were novel. In particular, the distinction drawn between *male/female* and *masculine/feminine* was made both to include non-binary speakers who disprefer the former, and to recognize binary trans men who prefer to present as *androgynous* or *feminine* and for whom this may be expressed in their voice. While identity is impossible to capture comprehensively for all speakers, if we are to take it into account in discussions of satisfaction, it would be a disservice to speakers to stop at *male*, *female*, *both*, and *neither*.

General Satisfaction

1. I dislike the sound of my voice.
2. I am happy with how my voice sounds.
3. I care about how my voice sounds.
4. I care about how my voice genders me to others.
5. I am afraid my voice will out me to others.
6. I avoid speaking aloud due to how my voice sounds.
7. I do not recognize my own voice in audio recordings.
8. I experience gender dysphoria surrounding my voice.
9. I sound like myself.

Internal Perception

10. My voice feels at odds with my gender identity.
11. My voice sounds masculine to me.
12. My voice sounds feminine to me.
13. My voice sounds androgynous to me.
14. My voice sounds male to me.
15. My voice sounds female to me.
16. I try to speak in a masculine way.
17. I try to speak in a feminine way.
18. I try to speak in an androgynous way.
19. I try to speak in a male way.
20. I try to speak in a female way.
21. I intentionally try to deepen my voice.

External Perception

22. My voice passes as a cis male's.
23. My voice is perceived as male by others.
24. My voice is perceived as female by others.
25. My voice is perceived as masculine by others.
26. My voice is perceived as feminine by others.
27. My voice is perceived as androgynous by others.
28. I am frequently misgendered over the phone.
29. I am frequently misgendered when speaking with others in person.
30. People tend to gender me correctly after hearing my voice.
31. People can tell I am trans based on my voice.
32. People assume I am not straight based on my voice.

Desire for Change

33. I wish my voice sounded more masculine.
34. I wish my voice sounded more feminine.
35. I wish my voice sounded more androgynous.
36. I wish my voice sounded more male.
37. I wish my voice sounded more female.
38. I do not want my voice to change significantly.
39. I am considering vocal therapy to change how my voice sounds.
40. I am considering vocal surgery to change how my voice sounds.
41. I want my voice to pass as a cis male's.
42. I want people to be able to tell I am trans based on my voice.
43. I want people to assume I am not straight based on my voice.

Participants completed this questionnaire during each session, ranking agreement with all statements on a five-point Likert scale (*strongly disagree* - *somewhat disagree* - *neither agree nor disagree* - *somewhat agree* - *strongly agree*), and could choose to skip any question. Responses were numerically coded from -2 (*strongly disagree*) to +2 (*strongly agree*). We expect overall satisfaction to increase with time on HRT, as well as both internal and external perception of one's voice as *male* or *masculine*, and an overall reduced desire for further change (Azul, 2015b,a; Nygren et al., 2009; Deuster et al., 2016; Hodges-Simeon et al., 2021). In the general satisfaction section, we predict **(1a)** an overall positive correlation with time on HRT and responses indicating positive opinions towards one's voice (Q2, Q9); **(1b)** a negative correlation with those indicating negative opinions (Q1, Q5-8); **(1c)** participant-specific change in responses regarding the focus or emphasis placed on the voice (Q3-4).

For internal perception, we predict **(2a)** that speakers' voices over time will feel less at odds (Q10) with their gender identity; **(2b)** that judgement of their voice sounding *male*, *masculine*, or *androgynous* (Q11, Q13-14) will increase over time; that **(2c)** judgment of sounding *feminine* (Q12, Q15) or *female* will decrease; **(2d)** that attempts to speak in a *masculine*, *male*, *androgynous*, or "deep" way (Q16, Q18-19, Q21) will decrease as the physiological effects of HRT reduce the need to behaviorally change one's speaking patterns in order to achieve desired vocal quality; **(2e)** that attempts to speak in a *feminine* (Q17) or *female* (Q20) way will be more variable between speakers, depending on identity and on their initial necessity of presenting as a cisgender woman pre-HRT.

In external perception, we similarly predict **(3a)** that perception of speakers' voices as *male*, *masculine*, or *androgynous* will increase over time (Q22-23, Q25, Q27) while **(3b)** perception of their voices as *female* or *feminine* (Q24, Q26) decreases; **(3c)** a negative correlation between time on HRT and misgendering both over the phone and in person (Q28-30); **(3d)** highly participant-specific change in whether individuals believe they are "clocked" as trans or assumed not to be straight (Q31-32) based on their voices; those who

pass as their desired gender pre-HRT may find their voice no longer suits them to others, while those who do *not* pass as their desired gender one year on HRT or for whom "passing" as a concept does not apply may find that otherwise.

We expect **(4a)** a decrease in overall desire for future change in the voice (Q33-38); **(4b)** a negative correlation between time and the desire for vocal therapy and surgery (Q39-40); **(4c)** participant-specific change in desire for one's voice to pass as that of a cis man, and for others to read one's voice as *trans* or *non-straight* (Q42-43), depending on individual circumstances regarding identity, passing, and concerns of physical safety.

Each question block's section combines the reporting of questionnaire results, and statistical analyses thereof, with qualitative discussion of participants' answers and interview responses in context. For each block, we first report overall change over time; the relationships between satisfaction responses and time on HRT were initially determined via a series of simple linear regressions ($\alpha=0.05$), collapsing all participant responses to provide a general overview across speakers. A mixed model incorporating speaker as a random effect was also used, to take individual variation into account. Both the correlation coefficient (r) between time and responses and the estimated fixed-effect coefficient for time (β) are provided.

Our next consideration is overall variation in participant responses. As each speaker had a maximum of twelve data points per question, with only five potential responses, session-to-session variation in individual speakers is often high; to that end, we focus here on participants' *first-* and *final-session* responses.¹ This does not directly take into account number of weeks on HRT, but allows us to introduce a more descriptive analysis of the experiences of individual participants or groups of participants, highlighting those who did not fit the overall trend as seen in the the linear regressions. In these sections, an "atypical" response

1. Though seven sessions of acoustic data are available for Marshmallow, only the first six sessions of data are available for the questionnaire results, as Marshmallow did not complete the questionnaire in their final session. Southern Boy chose not to respond to the satisfaction portion of the questionnaire during his first session; his second-session responses (at 8 weeks on HRT) are used here instead. The Q10 response for Wizard is similarly from their eleventh (penultimate) session, as this question was skipped in the twelfth.

defined as a positive or negative response to a given question that, for a given session, has a majority of negative or positive responses, respectively. "Ambivalent" responses ("neither agree nor disagree") are considered neither typical nor atypical, but are also discussed.

This examination of individual responses reveals underlying variation between speakers; while overall trends may be seen, a substantial number of participants for each question show little change, or change in a direction opposite than predicted. Here, numerous quotes are provided, giving context for their responses; participants after every questionnaire were asked if they had questions, comments, caveats, answers that they would like to elaborate upon, or uncertain interpretations of questions. Particularly given the nuanced and nebulous definitions of concepts such as "speaking in a male way", I argue that this commentary and discussion of the questions themselves provide as much insight as the responses to them, and that in questions where no trends across speakers are found, there is value in asking how and why participants differed in their answers.

6.2 Results: General Satisfaction

As expected, speakers' overall satisfaction with their voices did improve over time on HRT (Table 6.1), with speakers generally reporting decreased dislike (Q1) and increased happiness (Q2) with their voices, as well as sounding more like themselves (Q9). Time also correlated with a decrease in speakers' avoidance of speaking aloud (Q6) and experiences of vocal gender dysphoria (Q8); all of these results are as predicted. Overall, speakers did not change in the extent to which they cared about how their voice sounded or gendered them to others.

These results corroborate existing work on the effects of HRT on satisfaction, suggesting that overall satisfaction is improved over time on HRT. The lack of significant overall correlation with level of care about one's voice was also expected; speakers individually differed in this level of care and did not typically change in that level of care over time (see §6.2.1). Less expected was the lack of overall change in fear that one's voice would out one

#	General Satisfaction	r	β
1	I dislike the sound of my voice.	-0.392**	-0.027**
2	I am happy with how my voice sounds.	0.490**	0.033**
3	I care about how my voice sounds.	-0.064	-0.003
4	I care about how my voice genders me to others.	-0.033	-0.002
5	I am afraid my voice will out me to others.	-0.061	-0.005
6	I avoid speaking aloud due to how my voice sounds.	-0.307**	-0.018**
7	I do not recognize my own voice in audio recordings.	-0.101	-0.005
8	I experience gender dysphoria surrounding my voice.	-0.333**	-0.023**
9	I sound like myself.	0.394*	0.025**

Table 6.1: Correlation coefficients between "General Satisfaction" question results and time (r) and fixed effects estimate of time (β) all participants; higher r indicates stronger correlation, positive/negative r and β indicate direction of correlation. Significant effects are in bold; stars indicate level of significance (* $p \leq 0.05$, ** $p \leq 0.001$).

to others (Q5) and ability to recognize one's own voice in audio recordings (Q7). Though both of these questions were considered originally to be "negative" experiences, discussion with participants reveals nuance to their interpretation that likely contributes to the lack of significant change seen (§6.2.2).

Unlike Q5 and Q7, the lack of significant change in Q3-4 was anticipated, though for different reasons than seen; rather than high participant variation in the importance speaker's placed on their voice, there was near-universal agreement for Q3-4 throughout the study. Speakers overwhelmingly cared how their voice sounded and how it gendered them to others, with many in their first session claiming voice lowering to be the change that they were most looking forwards to, or even the primary reason they were on HRT (Southern Boy S1|W3).

One of the biggest reasons why I started testosterone is because I hated my voice, and like, I always felt it gave me away. Even if I kind of dressed masculine, and had an undercut, and had short hair, and did all the things to look masculine, I felt like when I opened my mouth people were like, "Oh, it's not a man, it's a masculine lesbian."
 And I was like, "No!" (Southern Boy S1|W3)

It must be recognized that responding positively to Q1, negatively to Q2, or otherwise indicating low vocal satisfaction in the questionnaire is an oversimplification of the complex-

ity of individual experiences; one speaker's "strongly disagree" is not the same as another's. Raine White, for example, responded typically for the entire general satisfaction question set (vocal dislike decreasing and happiness increasing over the course of the study, dysphoria reducing, etc.) and selected "strongly agree" for Q1 in Session 1. However, in the following interview, he discussed how this dislike of his voice was contextual (Raine White S1|W2).

I don't completely hate my voice. I think it's androgynous enough, I know it's not super high for a, you know, an assigned-female person. But like, I hate that – it doesn't sound bad to me, but I hate how it sounds to other people, and how they automatically assume that I'm female because of it. (Raine White S1|W2)

This is not a contradiction. Trans voices do not exist in a vacuum; Raine White did not dislike his voice intrinsically, but specifically how his voice was gendered by others. As such, we may anticipate satisfaction for him to be motivated largely by external perceptions, rather than personal feelings towards his voice. In contrast, other speakers focused on a more solitary understanding of their voice, with others' perceptions complementing but not informing their vocal goals. Gareth, who did not often socialize with others in person, expressed high vocal satisfaction pre-HRT ("I actually do like my voice. And I've worked to cultivate – not affect, but *cultivate* a voice that to me feels really grounded." -Gareth, S2|W1). In their words, voice change and the process of transition as a whole is in essence a spiritual journey (Gareth S1|W0). They were the only speaker to respond negatively to Q4 ("I care about how my voice genders me to others") in Session 1, and throughout the study, Gareth's reflections on their experiences were often focused internally, or on how their changing voice affected how they interacted with others and with the world (rather than inversely on how the world affected their voice).

I would say my goal is to be simultaneously more aware of my voice, and also to acclimate to my voice, and for my voice to be a vehicle conveying positive vibratory cues, to bring as much love and life onto this planet and it's a – this is a very sacred gift

we have, of speech. And to use that consciously as a way to uplift that vibration for myself, when in solitude, and to positively impact those around me in ways that may function at a very unconscious level. So, more specifically [*laugh*] I am not attached to achieving any particular – I don't have an image in my head of what I want my voice to be, and in fact I'm pretty happy with my voice right now. (Gareth S1|W0)

Thus, we may expect how Gareth sounds *to themselves* to be more indicative of satisfaction than how they sound to others. Though I use the terms "more" and "less indicative" here, the ability of these surveys to truly quantify satisfaction is already at question given the degree of nuance displayed. While I will discuss statistical tendencies and patterns seen, I primarily argue that work towards improving overall vocal satisfaction for transmasculine speakers on HRT must recognize these different perspectives, priorities, and framings of the voice and of the entire process of transition, and that these can all change over time.

6.2.1 *General satisfaction: Participant variation*

We see an overall trend towards higher satisfaction, though exceptions remain; all but three speakers expressed (via responses of "strongly" or "somewhat" agree) that they disliked the sound of their voice (Q1) in the first session. By the last, all but four disagreed with this sentiment (Figure 6.1). Atypical initial-session responses for Q1 ("I dislike the sound of my voice") reflect relatively higher initial satisfaction; these speakers were Gareth, Oz, and Renaud. In Q2 ("I am happy with how my voice sounds"), we see the same three, as well as Lizard and Southern Boy showing ambivalent ("neither agree nor disagree") responses.

The only atypical final-session response for either Q1 or Q2 was Marshmallow, who as said was able to complete only half of the study and thus in his final sessions often did not exhibit change in vocal characteristics to the extent of other participants (§4.3.1, §5.1.2); here, as with prior data, we will mark his final-session responses (and the initial-session responses of Oz) with †. Additionally, only three speakers (Kevin, Renaud, Wizard)

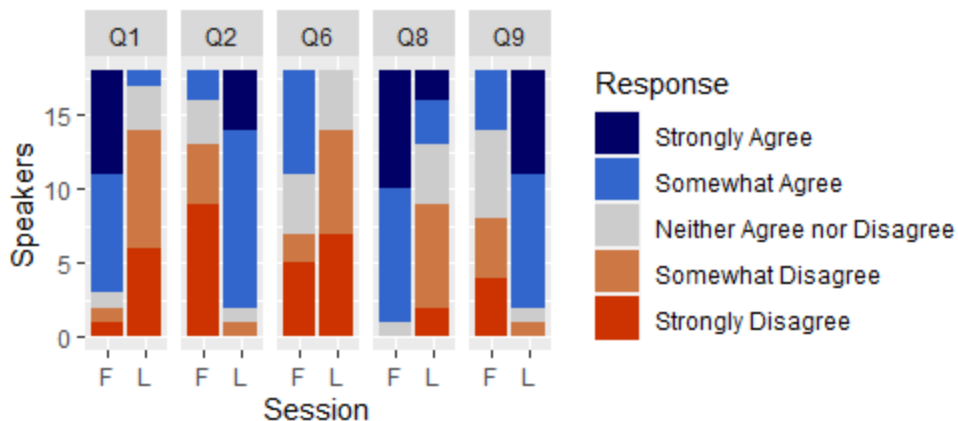


Figure 6.1: Distribution of (F)irst and (L)ast-session responses to all General Satisfaction questions exhibiting significant cross-participant change over time:

- Q1. I dislike the sound of my voice.
- Q2. I am happy with how my voice sounds.
- Q6. I avoid speaking aloud due to how my voice sounds.
- Q8. I experience gender dysphoria surrounding my voice.
- Q9. I sound like myself.

responded ambivalently to final-session Q1, and only one (Chimmy) for Q2; thus, save for Renaud, even those initially ambivalent ultimately showed increased satisfaction. However, high satisfaction by the end of the study does not always imply steady improvement of vocal well-being throughout; while the small number of data points per participant precludes rigorous statistical analysis, looking individually at how speakers' answers changed over time suggests that this increase in satisfaction is not always linear. Some speakers struggled with vocal discomfort, or took many months to experience noticeable change in their voice (as in those showing the "delayed descent" pattern in pitch, see §4.3.2).

Gareth's sensations of vocal "claustrophobia" (see §4.5), first expressed in Month 5, continued throughout the remainder of the study and exemplify the complicated emotional weight that can come from the changing voice. While the change was universally celebrated, for some there was also a period of mourning for the loss of the familiar (Gareth S10|W47). As described in §4.3.3, Gareth was one of three speakers (alongside ATM and Ty) to see a significant reduction of vocal range; their interviews exemplify how a speaker's quantitative

responses can conceal underlying change. They were also one of very few to be satisfied with their voice (and proud of what they had cultivated it into) pre-HRT.

I'm overall really happy with my voice. It's like, the most obvious physical change. But there's still this lingering sense of claustrophobia, of having lost the upper range. There's still some actual sorrow. Like, in an ideal world, I would want to have kept all of my upper range, especially when [...] you know, so many men sing in this upper falsetto range which I – I can't. So without some serious vocal training, that's gone. And – and I remain really curious about how my voice will continue to change.

(Gareth S10|W47)

It is because of situations such as these that I most strongly argue that accurately understanding the vocal satisfaction of trans speakers *requires* some amount of qualitative discussion. The questionnaire allows recognition of patterns, but out of context, it obscures the complexities of speakers' relationships with their voice; in measuring satisfaction numerically, we reduce Gareth's vocal transition from initial hard-won contentment, through claustrophobia, towards the joy of gender euphoria *in conjunction with* mourning, to the fact that they responded to Q1 with "strongly disagree" in both their first and final sessions.

For first-session Q6 ("I avoid speaking aloud due to how my voice sounds"), 7 speakers responded positively; by the last session, none did.² This trend of higher satisfaction was not without exception: of the four showing final ambivalence (Arrow, Gareth, Southern Boy, Wizard), two (Arrow, Gareth) had responded "strongly disagree" in the first session, suggesting for this pair an *increase* in avoidance of speaking aloud. This is particularly notable for Gareth, who did not speak to others often. Q8 ("I experience gender dysphoria surrounding my voice") is notable in that while overall improvement is seen, final results are more varied. In the first session, agreement was near-universal; all save for Renaud (ambivalent) expressed dysphoria surrounding the voice. In the final session, 9/18 participants

2. Additionally, of the four first-session ambivalent participants (ATM, Chimmy, Panda, Southern Boy), all except Southern Boy showed final-session disagreement.

	Ar.	AT.	Ch.	El.	Fr.	Ga.	Ke.	Kv.	Li.	Ma.	Oz.	Pa.	Ra.	Re.	Sa.	So.	Ty	Wi.	Summary
Q1						oo					o			x					(F) Low vocal dislike
Q1							x			•				x				x	(L) High vocal dislike
Q2						•			x		•			x		x			(F) High vocal happiness
Q2			x							o									(L) Low vocal happiness
Q6	x					x											x	x	(L) Ambi. to speaking aloud
Q8														x					(F) Ambi. to vocal dysphoria
Q8	••		•	x				x		•				x	x	•		••	(L) Persistent vocal dysphoria
Q9	x		x		•	•			x		x	x		•		•	x		(F) Sounding like oneself
Q9			x							o									(L) Not sounding like oneself

Table 6.2: Notable "General Satisfaction" Q1-9 responses by participant, in both (F)irst and (L)ast sessions. Bullets indicate response strength: oo *strongly disagree* | o *disagree* | x *neither agree nor disagree* | • *agree* | •• *strongly agree*.

disagreed, contrasting with the near-universal shift in Q1 and Q2. This reveals an important distinction: that "satisfaction" does not necessarily mean "complete cessation of dysphoria" (Wizard S11|W66)

I still definitely experience some dysphoria [...] either just listening to myself, or usually – where it gets me is if I'm having to speak loudly. I don't typically think that I need to [...] drop my voice because, you know, it's harder to make deeper sounds louder.

(Wizard S11|W66)

Table 6.2 summarizes participants' notable or atypical responses in both first and last sessions, for selected questions discussed here; i.e. those who still expressed lingering vocal dysphoria after the last session are given in Row 7.

Overall, while an overall significant *reduction* in dysphoria was seen, the rate of atypical responses reveals that vocal dysphoria is not eliminated, or reduced to a consistent extent, across all participants. Some (Elias, Kevin, Saint) showed final-session ambivalence after initial agreement, suggesting a reduction in dysphoria but possibly less so than other speakers. Renaud remained ambivalent throughout the study, reporting overall little vocal dysphoria and no change in this regard. Others, however, saw no such lessening at all; Arrow and Wizard in particular "strongly agreed" with Q8 in their final sessions, while Chimmy, Marshmallow[†], and Southern Boy "somewhat agreed". These speakers are notable for their

overlap with the atypical and ambivalent responders of Q1, Q2, and Q6 (see Table 6.2), suggesting a general association between the continued presence of gender dysphoria surrounding one's voice and a relatively lower improvement in satisfaction over the course of HRT. These responses may be compared to those of Q9 ("I sound like myself"), for which there was a broader initial spread; though 8/18 initially disagreed, several participants who otherwise reported that they did not like or were not happy with the sound of their voice responded atypically here. Those responding positively included Freddy, Gareth, Renaud, and Southern Boy; with the exception of Freddy, this group overlaps with those who otherwise expressed relatively higher initial-session satisfaction (Table 6.2), as we might expect. By the final session, we see near-universal agreement, with the exception of Chimmy (ambivalent) and Marshmallow[†].

Overall, we see general improvements to satisfaction, with a group of speakers (Arrow, Chimmy, Gareth, Kevin, Marshmallow[†], Renaud, Southern Boy, and Wizard) who expressed some level of general discontent surrounding their voices by the final session of the study.

6.2.2 *Mixed responses of Q5 and Q7*

Delving into speakers' explanations of their questionnaire responses, I found the phrasing Q5 "I am afraid my voice will out me to others" to be an error in methodology, as it has directly conflicting interpretations. Being "outed" (as a trans person) has different meanings depending on whether, and as what gender, one is attempting to pass.

A speaker who wishes to be read as a man in public may find their voice causes others to assume they are a woman. Alternatively, however, speakers closeted at work may worry that their changing voice will "clue in" their coworkers to their transition, which they are attempting to hide until they switch jobs. Thus individuals must consider the possibility and consequences of being "outed" - not only in the context of "failing to pass" as *male* (which time on HRT is likely to improve), but also situationally as *female* (which in that same

time is likely to become more difficult). There is the further consideration that "outing" as a concept is intrinsically associated with passing, which as discussed is a desired result for many, but not all, participants.

In the case of Q7, the phenomenon of not recognizing one's own voice in audio recordings is attested across various trans spaces. There are, however, multiple interpretations of this situation; on the one hand, pre-HRT individuals may find that their current voice in audio is unfamiliar or alien to them. This is the scenario initially intended by this question, in which over time on HRT, one's voice in recordings may become more familiar (Wizard S6|W26).

For the first time the other day, I heard my voice in a recording and wasn't extremely repulsed. So that's a new feeling! (Wizard S6|W26)

As with Q7, the question is ambiguous; it failed to specify *when* (relative to answering) this hypothetical audio recording would have taken place. The alternative interpretation is a related phenomenon: failure to recognize (or, relatedly, showing a dislike of) one's voice in recordings *after* some time on HRT. This is also attested by participants directly (Oz S8|W55). Unlike the initial interpretation, as speakers' own voices diverge from the pre-HRT baseline and become unfamiliar, this may be expected to *increase* over time on HRT; while speakers pre-HRT may find they do not sound as they feel they should, or as they sound in their head, their voices on HRT will likely not perfectly match that expected sound. This, combined with the rapid change and issues such as cracking and hoarseness, can create a sense of unfamiliarity in the voice that itself is a source of dysphoria and distress.

In regards to like, my voice causing me dysphoria, I've been keeping in touch with a friend over the summer, and we've been on voice call. Um, they don't use headphones, so sometimes if their mic is activated while I'm talking, I'll hear myself. And that's been very strange for me. [...] Bad, but I can't put my finger on why. Like, it's not that I don't like the sound of it, it's just so jarring. 'Cause it's definitely not what I sound like to myself. (Oz S8|W55)

6.3 Results: Internal Perception

The internal perception results corroborate prior findings that time on testosterone HRT correlates with a self-perceived increased *maleness* or *masculinity* of the voice, as well as an increase in perceived *androgyny*. We see an overall significant increase in (Q11) and (Q13-14) as predicted, as well as also the expected decrease in (Q12) and (Q15), indicating lower perceived *femininity* and *femaleness* of the voice, as well as in (Q10). There is, overall, a greater alignment of voice with sense of gender identity.

#	Internal Perception	r	β
10	My voice feels at odds with my gender identity.	-0.433**	-0.033**
11	My voice sounds masculine to me.	0.420**	0.029**
12	My voice sounds feminine to me.	-0.371**	-0.024**
13	My voice sounds androgynous to me.	0.232**	0.014**
14	My voice sounds male to me.	0.391**	0.027**
15	My voice sounds female to me.	-0.461**	-0.032**
16	I try to speak in a masculine way.	-0.073	< 0.002
17	I try to speak in a feminine way.	0.009	< 0.001
18	I try to speak in an androgynous way.	-0.027	-0.002
19	I try to speak in a male way.	-0.060	< 0.001
20	I try to speak in a female way.	-0.145*	-0.005*
21	I intentionally try to deepen my voice.	-0.118	-0.010**

Table 6.3: Correlation coefficients between "Internal Perception" question results and time (r) and fixed effects estimate of time (β) all participants; higher r indicates stronger correlation, positive/negative r and β indicate direction of correlation. Significant effects are in bold; stars indicate level of significance (* $p \leq 0.05$, ** $p \leq 0.001$).

The question of one's voice "feel[ing] at odds with [one's] gender identity" (Q10) shows stronger correlation with time ($r=-0.433$) as well as a stronger estimated coefficient ($\beta=-0.033$) than that of Q8 ($r=-0.333$, $\beta=-0.023$), a similar question that referred specifically to "experienc[ing] gender dysphoria surrounding [one's] voice". This is notable as it points to a distinction between physical dysphoria and discomfort; participants can find their voice feeling less "at odds" with their gender while still experiencing dysphoria surrounding it.

Unexpectedly, there is almost no overall correlation between time on HRT and reported attempts to speak in any particularly-gendered style (Q16-19), with the notable exception

of a slight decrease in trying to speak "in a female way" (Q20)³. The differences seen in responses between questions targeting voice *maleness/femaleness* and *masculinity/femininity* are also our evidence in support of drawing this distinction between these terms; we see a greater increase over time in describing one's voice as *masculine* than as *male*, and a greater decrease in *female* compared to *feminine*. Comparing the correlation strengths and slopes of the "My voice sounds X to me" question set, a hierarchy emerges; perception of one's voice over time grew, in descending order, *more masculine* > *more male* > *more androgynous*. Similarly, the negative correlation with time and one's voice sounding *female* was stronger than that between time and one's voice sounding *feminine*.

Finally, intentionally trying to deepen one's voice (Q21) trended slightly towards a negative correlation in the linear regression, but was ultimately insignificant ($p=0.090$) while the mixed effects model did show a significant effect of time ($p<0.001$).

In examining these responses, we must fundamentally acknowledge that no two participants will have the same definition or criteria for "sounding" in a particular way. We have previously discussed acoustic correlations of perceived gender, masculinity and femininity in the voice, and the extent to which trans individuals specifically tend to be aware of these features; however, awareness of acoustics may do very little to affect an experience of dysphoria or disconnect. Sounding *male* is an inherently subjective concept; for some speakers, it may mean vocally resembling a male relative or male acquaintances, displaying typical or stereotypical male vocal characteristics (e.g. low pitch, monotonous speech), passing as a cis man in public, or simply "sounding like oneself" while knowing that this "self" is male. Even if tracking change in pitch through an application shows an objective change in the voice, subjective perception (by oneself or by others) may differ.

3. Results for Q20 were only just significant ($p=0.05$ precisely, in contrast to all other significant results for which $p\leq 0.001$).

6.3.1 Internal Perception: Participant variation

Responses to Q10 ("My voice feels at odds with my gender identity") parallel those of Q8, being a similar sentiment but specifically referencing identity rather than dysphoria. Most participants agreed with this statement in their first sessions, with ATM as the only atypical response and Gareth and Renaud showing initial ambivalence; this conforms with the relatively high pre-HRT satisfaction also shown by the latter two.

By the final session, while most speakers disagreed with Q10, four (Arrow, ATM, Kevin, Southern Boy) expressed ambivalence and three still showed agreement (Chimmy, Kel, Marshmallow[†]). Of those four, all but ATM did experience a decrease in how "at odds" their voice was with their identity, but possibly not to the same extent as others; Arrow and Southern Boy in particular also had expressed lingering dysphoria and final-session ambivalence towards speaking aloud, adding to an overall pattern of lower satisfaction.

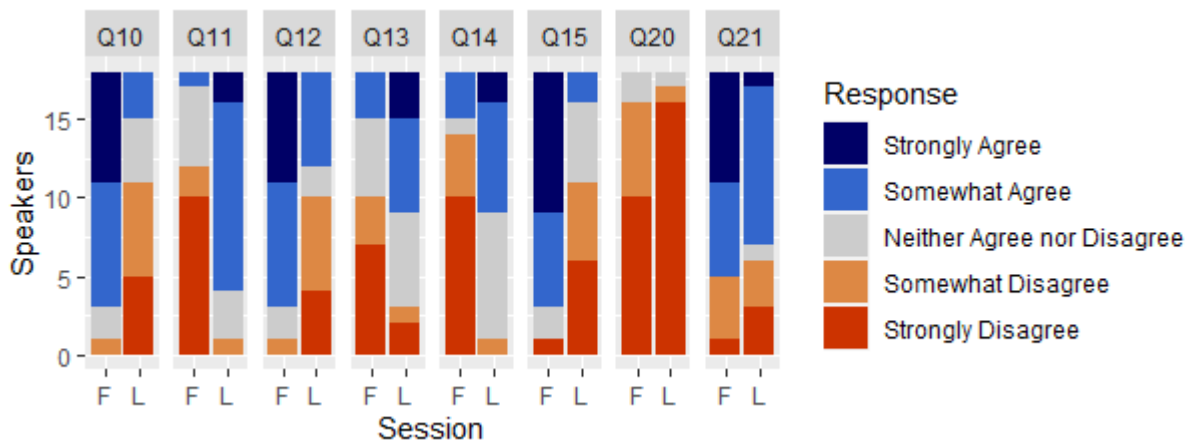


Figure 6.2: Distribution of (F)irst and (L)ast-session responses to all Internal Perception questions exhibiting significant cross-participant change over time:

- Q10. My voice feels at odds with my gender identity.
- Q11. My voice sounds masculine to me.
- Q12. My voice sounds feminine to me.
- Q13. My voice sounds androgynous to me.
- Q14. My voice sounds male to me.
- Q15. My voice sounds female to me.
- Q20. I try to speak in a female way.
- Q21. I intentionally try to deepen my voice.

ATM, notably, went from disagreement to ambivalence⁴, suggesting that their voice grew *more* at odds with their gender identity over time. Like Arrow and Southern Boy, Chimmy reported continued dysphoria (Q8), and this appears to be an extension of that.

Variation in vocal maleness and masculinity

We now examine the "My voice sounds..." questions, beginning with the pair Q11 *masculine* and Q14 *male*. The results here exemplify the importance of considering individual speaker goals in determining whether and to what extent vocal change on HRT has "succeeded". In the first session, more agreed with *male* (Gareth, Oz[†], Southern Boy) than "masculine" (Oz[†]), while more were initially ambivalent to *masculine* (ATM, Arrow, Gareth, Renaud, Southern Boy) than to *male* (ATM). Given Oz's time on HRT, it is in fact the case that *no* participant pre-HRT believed their voice sounded *masculine*, including those who agreed it sounded *male*.

Though we may initially suspect this to be influenced by the previously-described "identity-first" reasoning of vocal gender ("I am male, therefore my voice is a male one, even if it is not masculine") (see §2.3.1), of these three speakers only Southern Boy indicated in Session 1 that they would use the term *male*; by Session 12, he responded ambivalently to that term. Additionally, while Gareth and Southern Boy expressed "somewhat" likelihood to call themselves *masculine*, Oz strongly disagreed with both this term and with *male* throughout the study (see §3.4.2). Ergo, we see here three speakers who are not precisely *male*, and did not view their voice as initially sounding *masculine*, but *did* view it as initially sounding *male*. Furthermore, there is some overlap between this group and those who are more satisfied with their pre-study voice, as all of Gareth, Oz, and Southern Boy had expressed relatively high satisfaction in their first sessions.⁵

4. Examining month-to-month change shows ambivalence for ATM's final three sessions, preceded by near-exclusive disagreement.

5. As did Renaud, who continues his pattern of ambivalence.

By the final session, these responses had shifted drastically; 14/18 speakers agreed that their voice sounded *masculine* (Q11), and 9/18 agreed it sounded *male* (Q14). There was only a single disagreement for both questions; Kevin for Q11, and Arrow for Q14 (see Table 6.4). Of the four participants who did *not* indicate vocal masculinity in the final session (Chimmy, Kevin, Oz, Ty), before assuming this is a universal indicator of low satisfaction, we must first ask whether vocal masculinity for these participants was in fact a goal; the experiences described by each participant in interviews were tremendously different.

Chimmy and Kevin had previously indicated lower final satisfaction, responding atypically or ambivalently to questions about voice-gender mismatch and dysphoria. We may thus presume this lack of perceived vocal masculinity to be a contributing factor; in the case of Kevin, who explicitly referenced wishing to sound like a cisgender man, this likely is the case. For Chimmy, this is more complicated; in the final-session identity questionnaire responses, Chimmy was one of only two participants "very unlikely" to use *masculine* for himself. Consistently throughout the study (§3.4.2), the *only* word in the identity questionnaire that he indicated he would use for himself is *queer*. Similarly, Chimmy's interviews indicate a goal better-described not as reaching towards vocal masculinity, but away from vocal femininity (Chimmy S1|W7).

I want to not sound unmistakably and instantly identifiable as female. Even if someone just isn't sure how to gender me, that would be a huge improvement. That would be a starting point! (Chimmy S1|W7)

However, being gendered as female in public was nonetheless a source of major distress, to which vocal perception as *not-masculine* is a contributor. Chimmy describes largely external motivations (i.e. social dysphoria) for changing the way his voice is perceived; in general, his vocal dissatisfaction is frequently framed as relating to others' perceptions of his voice (particularly in the context of his work, which involves numerous interactions with strangers over the phone), de-emphasizing self-perception. This is not to say that he does not consider

	Ar.	AT.	Ch.	El.	Fr.	Ga.	Ke.	Kv.	Li.	Ma.	Oz	Pa.	Ra.	Re.	Sa.	So.	Ty	Wi.	Summary
Q10		o				x								x					(F) Voice less "at odds"
Q10	x	x	•				•	x		•						x			(L) Voice still "at odds"
Q11	x	x				x					•			x		x			(F) High vocal masculinity
Q11			x					o			x						x		(L) Low vocal masculinity
Q14		x				•					•					•			(F) High vocal maleness
Q14	o		x					x	x		x	x				x	x	x	(L) Low vocal maleness

Table 6.4: Notable "Internal Perception" Q10-14 responses by participant, in both (F)irst and (L)ast sessions. Bullets indicate response strength: oo *strongly disagree* | o *disagree* | x *neither agree nor disagree* | • *agree* | •• *strongly agree*.

how he sounds *to himself* to be important, only that for some speakers, internal perception may overall be a poorer predictor of satisfaction compared to external; changing how one personally views one's voice may do little to solve the issue of being misgendered in public.

In contrast, despite also not reporting an increased sense of vocal masculinity, Ty and Oz have shown extremely high overall increases to satisfaction, corroborating this in interviews. Ty began by "strongly disagreeing" that his voice sounded *masculine*, across the first six sessions; over the last six, this changed to "somewhat disagree" and then ambivalence. As such, Ty shows the same typical response pattern seen by the speakers who perceived their voices changing from *not-masculine* to *masculine* (such as Arrow, Elias, and Lizard); it is only that his "endpoint" was ambivalence.

Oz's responses to Q11 were the reverse, responding with a steady "somewhat agree" for the first six sessions that in the latter six began to change, with a "strongly" disagree in S10 and a "somewhat" in S11; over time, we find Oz grew to perceive their voice as *less masculine*, though their overall satisfaction with their voice improved. This relates to the shifting of identities and presentation over the course of HRT discussed in 3, an experience which Oz described in detail; as they grew more comfortable in their identity as non-binary, changes in perceived vocal masculinity and satisfaction were not correlated.

The patterns seen for the corresponding Q14 tell a similar story; though nine speakers did not agree that their voice after the first year sounded male, their responses and reasoning differed. Only one (Arrow) outright disagreed; of the eight others, only three (Kevin, Panda,

and Ty) in that same session also reported *identifying as* male. Even among those three, though Kevin and Ty both expressed desire for a cis-passing voice, Panda described their vocal goals "more androgynous, but masculine-leaning in that androgyny" (Panda S1|W0). The remaining five ambivalent responders (Chimmy, Lizard, Oz, Southern Boy, Wizard), as well as Arrow, did not use or were also ambivalent to the term *male* in reference to themselves during the final session. For several of this group, particularly Oz (who showed high general satisfaction), self-perception of vocal maleness may in fact *not be a useful indicator* of satisfaction; they are not male, and do not necessarily desire to sound male. For the others, relatively lower satisfaction has still been seen in prior responses, suggesting that despite not identifying as male, *male-ness* of the voice can still be an indicator of satisfaction.

Variation in androgyny

Responses to Q13 in both initial and final sessions were comparable to those of Q11 and Q14, suggesting an overall increase in perceived vocal androgyny. Three speakers (Chimmy, Gareth, Raine White) responded atypically (indicating high androgyny) in Session 1 (see Table 6.5); notably, of these, only Gareth also expressed high initial vocal maleness or masculinity. Through perceiving their pre-HRT voices as *androgynous*, both Chimmy nor Raine White "strongly disagreed" that it was *masculine* (and "slightly disagreed" with *male*).

By the final session, a plurality of speakers did report vocal androgyny; most who were ambivalent typically either initially disagreed (Lizard, Panda, Saint), thus showing this same pattern of increased androgyny if to a lesser extent, or were initially ambivalent as well (Oz, Renaud). This matches the overall correlation found between time on HRT and perceived androgyny, alongside maleness and masculinity. However, there are a few speakers for whom a change in vocal androgyny over the course of the study did not have a parallel change in maleness or masculinity. The three atypical Q13 Session 1 responders also responded atypically or ambivalently in Session 12 (Table 6.5); all who showed relatively high perceived

	<i>Ar.</i>	<i>AT.</i>	<i>Ch.</i>	<i>El.</i>	<i>Fr.</i>	<i>Ga.</i>	<i>Ke.</i>	<i>Kv.</i>	<i>Li.</i>	<i>Ma.</i>	<i>Oz</i>	<i>Pa.</i>	<i>Ra.</i>	<i>Re.</i>	<i>Sa.</i>	<i>So.</i>	<i>Ty</i>	<i>Wi.</i>	Summary
Q13		x	•			•					•		•	x		x		x	(F) High vocal androgyny
Q13			x		oo	o			x		x	x	oo	x	x				(L) Low vocal androgyny

Table 6.5: Notable "Internal Perception" Q13 responses by participant, in both (F)irst and (L)ast sessions. Bullets indicate response strength: oo *strongly disagree* | o *disagree* | x *neither agree nor disagree* | • *agree* | •• *strongly agree*.

vocal androgyny then indicated relatively low vocal androgyny by the end of the study, a decrease contrasting with the overall pattern seen. The only other atypical responder in Session 12, Freddy, both began and ended with low perceived vocal androgyny.

This group exhibits a disconnect between reported vocal androgyny and maleness or masculinity. Chimmy moved towards ambivalence for all three by the final session, Gareth saw a sharp decrease in perceived androgyny but very little change in maleness and masculinity, Raine White saw a similar decrease alongside an increase in maleness/masculinity, and Freddy reported consistent low vocal androgyny despite an increase in maleness/masculinity.⁶ The disconnect is notable because for most individuals who have not yet experienced testosterone's effects, the processes of shifting one's voice *towards* androgyny versus towards maleness are overlapping, if not entirely identical to a point. That Q13 generally patterns with Q11 and Q14 corroborates this; if we define androgyny as some intermediate area in which a voice is not immediately identifiable as *male* or *female*, then change away from a female-sounding voice necessarily becomes both more male-sounding and more androgynous.

This has two caveats: First, even this *female* → *androgynous* → *male* framing allows for a voice already viewed as androgynous to, over the course of HRT, move *away* from androgyny but still towards greater maleness/masculinity (as potentially exemplified in Raine White). More importantly, as discussed in §2.1, this linear mapping and the corresponding idea

6. Note that Freddy and Raine White have not responded atypically or ambivalently to any other significant question previously examined (save for Freddy "sounding like [him]self" pre-HRT, Q9); they both show an increase in overall satisfaction comparable to speakers for whom androgyny and maleness/masculinity are linked.

that *lower pitch*⁷ = *more male* is only one framing, based on the bimodal distribution of various characteristics of human voices that differ between genders. In this way it reflects the binary model of human gender itself; vocal androgyny, like non-binary identity, is framed as inherently intermediary between the two "male" and "female" means. While this may be the case in production for any given acoustic property, it does not follow that speakers will necessarily *perceive* vocal androgyny as such; they may not correlate androgyny with maleness or masculinity, but view it as a separate axis upon which change may occur. The responses of Freddy in particular do not match a *female* → *androgynous* → *male* framing, as he maintained a "strongly disagree" response for perceived vocal androgyny consistently across all twelve sessions, despite increasing in maleness and masculinity.

Variation in femaleness, femininity, and mixed-gender perception

In Session 1, speakers near-universally (15/18) agreed that their voice sounded both *feminine* (Q12) and *female* (Q15); only Oz[†] responded atypically for both. Chimmy and Renaud were the only ambivalent responses to initial Q12 (i.e. reporting that their voice to them sounded *female* but not *feminine*), ATM and Southern Boy for Q15 (vice versa). All four also expressed initial ambivalence or agreement with vocal androgyny, and there was substantial overlap (save for Chimmy) with those reporting initial vocal masculinity and maleness.

The overall trend towards decreased femaleness/femininity of the voice over time holds true for most speakers, though with a substantial number of exceptions, and final-session responses show a clear distinction made between femaleness and femininity for some speakers. Six (Arrow, ATM, Kel, Lizard, Marshmallow[†], and Ty) responded atypically for Q12 in their final sessions, and two (Kel and Marshmallow[†]) for Q15. Kel and Marshmallow[†] were the only participants to report that their voice sounded both *feminine* and *female* to them by the end of the study; none responded atypically to *female* exclusively, though Southern Boy,

7. Or lower formants, longer VTL, etc.

	Ar.	AT.	Ch.	El.	Fr.	Ga.	Ke.	Kv.	Li.	Ma.	Oz.	Pa.	Ra.	Re.	Sa.	So.	Ty	Wi.	Summary
Q12			x								o			x					(F) Low vocal femininity
Q12	•	•		x			•	x	•	•								••	(L) High vocal femininity
Q15		x									oo						x		(F) Low vocal femaleness
Q15	x						•	x		•	x						x	x	(L) High vocal femaleness

Table 6.6: Notable "Internal Perception" Q12-15 responses by participant, in both (F)irst and (L)ast sessions. Bullets indicate response strength: oo *strongly disagree* | o *disagree* | x *neither agree nor disagree* | • *agree* | •• *strongly agree*.

Oz, and Wizard did show exclusive ambivalence to *female* (see Table 6.6). Notably, both Kel and Marshmallow[†] described what I refer to here as "mixed-gender" vocal perception, responding typically for Q11, Q13, and Q14 in their final sessions and thus also reporting increased vocal maleness, masculinity, and androgyny. This suggests, as discussed in §6.3.1, the need for an understanding of vocal gender perception that does not place male and female as opposed, with androgynous in the middle. Rather, these speakers' impressions of their voices were at once *male*, *masculine*, *androgynous*, *female*, and *feminine*.

In the case of Marshmallow, we might anticipate this as the result of having completed only six sessions (27 weeks), such that mixed-gender perception represents still-ongoing change. Examining the responses of other participants closest to the 27-week mark, we do indeed see several expressing mixed-gender perception (ATM S6|W27, all but *female*; Freddy S8|W30, all but *androgynous*; Kevin S6|W26, all but *feminine*; Saint S6|W25, all). This does suggest that Marshmallow's final-session identification of his voice as *female* in particular may have been in part the result of this early cutoff, as Saint, Kevin, and Freddy would go on to no longer perceive their voice as such. However, not only do we still see Kel reporting their voice as sounding *female* after twelve full sessions, but other participants who did not agree with *female* nonetheless exhibited mixed-gender vocal perception in other ways. Apart from several expressing some combination of *masculine*, *male*, and *androgynous*, we see more varied combinations as well: Arrow with *masculine*, *feminine*, and *androgynous*; ATM with all but *female*, Ty with *feminine* and *androgynous*, and notably Lizard with both *masculine* and *feminine* (but neither *androgynous* nor *male/female*).

Thus we see a total of six participants who by the final session identified their voice as "sounding" like some combination of genders or identities that included *feminine/female* alongside *male/masculine/androgynous*. While this is effectively a non-binary perception of the voice, it does not necessarily correspond to non-binary identity; though Arrow, ATM, Kel, and Ty did use the term *non-binary* in reference to themselves (with all but Ty also using *genderqueer* and Arrow using *genderfluid*, see §3.4.2), Lizard and Marshmallow did not.⁸ Further, while this does suggest mixed-gender vocal perception to be more *characteristic* of non-binary speakers, other non-binary participants (Elias, Gareth, Kevin, Oz, Panda, Southern Boy, and Wizard) showed final-session vocal perception only in the male/masculine/androgynous combination. The experiences of these speakers above all support the idea that *male/female* and *masculine/feminine* are not linearly opposed in perception of the voice, any more than they are in gender identity and presentation.

Just as there was some overlap between those with higher general satisfaction and those with more male, masculine, or androgynous-perceived voices, there is also some overlap between lower general satisfaction and perceived vocal femininity/femaleness. We see Arrow, Kevin, and Marshmallow all expressing this pattern; often, speakers clarified that perceived femininity was situational, in specific contexts or around certain people (Arrow S10|W58). However, this does not appear reflective of an overall trend. Kel, despite reporting both high *femaleness* and *femininity*, notably did *not* respond atypically in Q1-9, nor did Lizard and Ty despite high reported *femininity*, nor Oz with ambivalence towards *female*.

At the beginning of the year, it was like, this is where your comfort level actually is for your voice, stop straining [...] and now I'm finally getting to where I'm regularly keeping my voice in that area, with strangers or other people. But it still – with my mom, it still goes up. I don't really know why. (Arrow S10|W58)

8. In their final session, neither explicitly claimed to be *binary* either, and Lizard throughout most of the study expressed greater likelihood of describing himself as *female* than as *male*; however, he also preferred terms such as *FTM* or *transsexual* over *transgender*, suggesting that use of *female* in this case may be seen in a biological or medical-focused context rather than one of identity.

Similarly, not all those with lower general satisfaction (particularly as identified through Q8 regarding dysphoria) also identified their voices as *feminine/female*, with Chimmy, Southern Boy, and Wizard reporting lingering vocal dysphoria at the conclusion of the study but none of the three believing their voice to sound *feminine* (and the latter pair only showing ambivalence towards *female*). This illustrates a disconnect between physical dysphoria surrounding one's voice and the way that voice is gendered; a transmasculine speaker on HRT is not necessarily freed of vocal dysphoria just because they no longer find they sound *female*.

6.3.2 *Participant variation and what it means to speak in a "way"*

There was for the most part no overall significant change in the extent to which speakers actively tried to speak in a particularly-gendered way; however, the idea of a "male way" or "masculine way" to speak is an inherently nebulous one. The use of this phrasing in the questionnaire was less to measure the use of specifically gendered speech patterns and more to provoke discussion, in the subsequent interviews, about how participants defined and interpreted these gendered speech styles. This wording captured a wide range of possible individual definitions for what a male/masculine/etc. "way" of speaking actually entailed, and as much value may be gained from participants' discussions of these definitions as from their quantitative responses themselves.

For the most part, the definitions given for a masculine or male "way" of speaking largely corresponded with the salient aspects of vocal gender described in §2.1, such as lower pitch, as well as the more stereotyped reduced pitch variation ("speaking in monotone"). Feminine or female ways of speaking were similarly based on inflection, particularly the use of high rising terminal ("uptalk"), use of creaky voice ("vocal fry")⁹, and or high pitch variation ("lots of

9. As described briefly in §2.2.5, creaky voice may have a lowering effect on perceived F0 and thus is occasionally suggested in transmasculine spaces as a method of lowering pitch before or during HRT. Its effects on gender perception in the voice, and use in the trans community, are to be the topic of future work, but it is worth noting here that though the majority who mentioned it drew an association with the speech of young women, more than one participant did reference vocal fry as characteristic of *male* speech.

ups and downs"). Describing female/feminine-coded speech as "musical" was also common, with several participants with singing experience drawing explicit connections between their experiences singing (typically female choir parts, soprano or alto) and perceived femininity in their voice (ATM S1|W1).

I've been a singer since I was really little. [...] Everyone always told me, "Your voice is so beautiful!" And it's always been, "Thank you, but there's something wrong with it." [...] Everyone says, "You have a very melodic way of speaking." And that codes feminine, and I hate it. (ATM S1|W1)

Not all participants drew distinctions between male/female and masculine/feminine, but those who did typically focused on pitch versus inflection. Ty, for example, initially described his own voice as having feminine inflectional qualities due to his non-binary identity, but wanted his voice to be otherwise unambiguously male (Ty S1|W10).

I think overall I'm kind of hoping for more of a cis male voice. [...] Since I kind of identify more on the non-binary line, I kind of have a lot more feminine qualities as far as, like, speech articulation and stuff. [...] I tend to speak with more of a lilt to it, so my voice tends to go off [higher], or I kind of end my sentences on a higher note. (Ty S1|W10)

Near the end of the study, Ty's voice continued to show these qualities (Ty S11|W62):

I kind of fall under more, like, a feminine-masculine energy. So I feel like my voice kind of reflects that, a little bit. Because I'm not so worried about trying to sound, like, super deep or masculine. (Ty S11|W62)

That said, this lack of statistical significance was somewhat unexpected; regardless of how individuals defined a gendered "way" of speaking, the phrase "try to speak" refers to the use of behavioral techniques to alter the voice, and it was predicted that speakers

who initially used voice masculinization techniques in particular would do so less as HRT "masculinized" the voice. Such a prediction matches the "expected transmasculine vocal narrative" described in §1 (the process of going on HRT, seeing a sharp drop in pitch, and reaching a cis-male-typical vocal range with no need for further training or intervention and no issue passing vocally as male), which as described has come under recent scrutiny (e.g. from Azul (2015b)). This prediction appears inaccurate; those who initially expressed attempts to speak in a particular "way" generally maintained these attempts throughout the study, and those who initially worked to deepen their voice generally continued to do so.

There are some exceptions. All participants responded affirmatively or ambivalently to Q16 ("I try to speak in a masculine way") in the first session; Oz and Lizard alone responded negatively in the final session, meaning that only these two participants reduced efforts to sound *masculine* over time. In contrast, we see Ty and Wizard move from ambivalence to agreement (Wizard in particular "strongly agreeing"), suggesting increased efforts; as Wizard reported persistent vocal dysphoria by the final session (and Ty ambivalence towards both maleness and masculinity), we may view these increased efforts as a response to the majority of vocal changes from HRT having occurred. Wizard addressed this in the penultimate session, deciding that now that they've seen most change they can expect, they will focus on practicing avoiding instinctive use of a higher pitch (Wizard S11|W66).

I feel like I was the most happy with it when I was noticing, like, the big changes. [...] I would like my voice to continue dropping, but I think at this point maybe, unless another year or so of T is just gonna really continue to drop it down — which I know does happen for some people — I think at this point it's mostly training myself to remember to not [...] heighten my voice in stressful situations, and things like that.

(Wizard S11|W66)

Again, we see a distinction between *masculine* (Q16) and *male* (Q19); Panda and Raine White did not, in Session 1, claim to intentionally speak in a *male* way. By the final session,

disagreement here was more common, seen in ATM, Lizard, Oz, and Panda (with Raine White now ambivalent). More participants reduced their efforts to sound specifically *male*, but not enough to form any discernible trend. Chimmy and Wizard here instead showed an increase, moving from ambivalence to agreement (again "strongly" in Wizard's case); once more, this may reflect the continued post-study dysphoria expressed by both.

As for speaking in an *androgynous* way (Q18), responses here were extremely varied. Most participants who initially agreed (Arrow, Elias, Lemon) or disagreed (Freddy) continued to do so, many (ATM, Gareth, Kel, Lizard, Oz[†], Raine White, Renaud, Ty) saw a decrease from agreement/ambivalence to ambivalence/disagreement, while others saw an increase (Marshmallow[†], Panda). However, examining month-to-month responses for these participants reveals high variation and little discernible pattern. The opposite is true for *feminine* (Q17), where the lack of significant change is the result of little variation at all; initial responses were overwhelmingly negative with the sole exception of Lizard, and final responses were much the same with the exception of Elias.

The lack of change in *male/masculine/androgynous* contrasts with Q21 ("I intentionally try to deepen my voice"), which did see a negative effect of time on HRT in the mixed effects model (though no correlation in the simple linear regression). This is reflected in individual responses, where most participants saw either no change or a decrease in this intentional deepening. Despite this, 11/18 participants still reported intentional voice deepening in their final sessions, though all but Wizard "somewhat" rather than "strongly" agreed with Q21 (contrasting with 6/18 "somewhat" and 7/18 "strongly" agreeing responses in Session 1). Thus, while we see overall reduction in voice lowering efforts, such efforts *are still present* across speakers. Chimmy and Wizard retained agreement through Session 12, corresponding to this pair also being the only speakers to show increased efforts to "speak in a male way" (Q19). ATM and Freddy both initially disagreed with Q21, but agreed in their final sessions, suggesting a desire to intentionally deepen their voice which developed over time.

That intent to deepen one's voice shows significant reduction while the broader "try to speak in a male/masculine/androgynous" way does not is notable; as association of a deeper voice with greater *maleness*, *masculinity*, and/or *androgyny* (for those with voices initially of a cis-female-typical pitch) would suggest that these should change in parallel. That we do not see this may be further evidence for a decoupling of acoustic factors from their gendered associations (i.e. for these speakers, a deepening of the voice is not necessarily also a masculinization of the voice). More likely, however, is that because speaking in a particular "way" simply encompasses more behavioral change than only voice deepening, those who work to deepen their voice without also adopting other gendered behavioral characteristics of the voice (e.g. inflectional patterns) may report the latter rather than the former.

Finally, we see no significant change in Q17 ("I try to speak in a feminine way"), but a trend towards increased disagreement for Q20 ("I try to speak in a female way"). Nearly all participants disagreed with both questions initially; the downwards trend was not the result of participants shifting from agreement to disagreement, but overwhelmingly from "somewhat" to "strongly" disagree. Lizard was the only atypical responder for the former; he describes his intent in earlier sessions to adopt a more feminine vocal style to avoid confrontation in social interactions (S1|W4), a sentiment also expressed by other speakers.

If I'm not sure what kind of environment I'm in, I'll make my voice sound more feminine – it's kind of like, maybe it's better to play into people's assumptions. [...] I don't really like to correct people that much. I almost make it easier for everyone involved. It's like, let's get this over with, whatever. (Lizard S1|W4)

By the final session, Lizard no longer agreed with either question, nor did any other participants with the exception of Elias for Q17. Ty's case is also notable here, as he had situations in his daily life in which he was not out as trans, and thus continued to need to pass as female in certain circumstances. He describes noticing himself speaking in a more feminine way, but being uncertain of the directionality:

I sometimes catch myself going higher, or whatever, than I would normally. So I don't know if it's like, "Oh, I'm being perceived as female so I need to instinctively sound more feminine", um, or if it's just me being — like, noticing that kind of speech more, just because of the situation. (Ty S11|W62)

Overall, although individual participants' responses can inform us as to their overall experiences, change in efforts to speak in a particular "way" are so participant-specific, and so dependent on individuals' definitions of such ways of speaking, as to not result in any clear trends. The exception is in Q20 ("I try to speak in a female way") which as said showed an unexpected significant negative correlation with time on HRT. This change was not one from agreement to disagreement overall, but rather an increase in *strength* of agreement; from 10/18 "strongly disagree" in the first session to 16/18 in the last.

Examination of participants' discussions of this question, and of their lives, reveals the likely reason for this: over the course of the first year of medical transition, many speakers were socially transitioning as well, which included coming out to friends, family, and coworkers. For those who did so, trying to "speak in a female way" was reflective of trying to effectively pass as a cisgender female in the family or workspace, to not let on that one was intending to transition until that process was already underway, for reasons of safety, social discomfort, or fears of discrimination. Over the course of the first year of HRT, these speakers no longer tried to "put on" this female speech style.

6.4 Results: External Perception

As expected, we see an overall positive correlation between time on HRT and participants' reports that their voices were perceived as more *male*, *masculine*, or *androgynous* by others (Q22-23, Q25, Q27), as well as a negative correlation with perception as *female* or *feminine* (Q25, Q26). There was also the predicted overall decrease in tendency to be misgendered both on the phone and in-person (Q28-29) and an increase in being correctly gendered.

Notably, both being read as *trans* and as *not straight* based on one’s voice (Q31-32) were predicted to show no overall change due to variability in speaker situation (and interpretation of the question); instead, like with Q21, the linear regression for both approached but did not quite (or, in latter case, barely) reach significance, while the mixed effects models for both did show an extremely small but significant effect of time on HRT.

#	External Perception	r	β
22	My voice passes as a cis male’s.	0.448**	0.033**
23	My voice is perceived as male by others.	0.531**	0.039**
24	My voice is perceived as female by others.	-0.432	-0.030**
25	My voice is perceived as masculine by others.	0.486**	0.033**
26	My voice is perceived as feminine by others.	-0.391**	-0.026**
27	My voice is perceived as androgynous by others.	0.309**	0.019**
28	I am frequency misgendered over the phone.	-0.302**	-0.030**
29	I am frequency misgendered when speaking with others in person.	-0.334**	-0.027**
30	People tend to gender me correctly after hearing my voice.	0.407**	0.036**
31	People can tell I am trans based on my voice.	0.126	0.010
32	People assume I am not straight based on my voice.	0.137*	0.009**

Table 6.7: Correlation coefficients between "External Perception" question results and time (r) and fixed effects estimate of time (β) all participants; higher r indicates stronger correlation, positive/negative r and β indicate direction of correlation. Significant effects are in bold; stars indicate level of significance (* $p \leq 0.05$, ** $p \leq 0.001$).

As with the previous question set, we see a hierarchy in external gendered perception of one’s voice based on the correlation strength and fixed effects coefficient estimate: *more male > more masculine > more androgynous*. Notably, this is not the same ordering as in the internal perception set, where there was a stronger increased of perceived masculinity as opposed to maleness. There are several explanatory factors to consider here; first, external perception is based entirely on speakers’ *reports* of how others perceive them. Without a corresponding perceptual study, we cannot be certain how speakers’ voices are in fact gendered by strangers, and it may be that the balance of *male/masculine* is skewed.¹⁰ This

10. The lack of direct confirmation of external perception does also mean that speakers’ intuitions may not be entirely correct; e.g. speakers may believe they are read as male/female more often than they actually are. However, I argue that when it comes purely to understanding vocal satisfaction, how one *believes* one’s voice is perceived will provide more insight into how happy one is with one’s voice than how it is *actually* perceived. A trans man who believes his voice is read as female by others will likely be dissatisfied, regardless of how accurate that belief is.

is related to the findings of Quinn et al. (2021) that trans speakers' satisfaction with their voices tended to correlate more with self-perception of vocal gender than with external perception. Even to those who like or ambivalent about their voices internally, and for whom external perception is more relevant, the inability to know for certain how others read one's voice make it a difficult means of gauging "success" (if being read a certain way by others is that speaker's personal goal).

In discussing the questionnaire and how they went about answering various questions, speakers described basing their responses to Q21-Q27 on strangers' comments and forms of address, including unprompted use of pronouns and gendered terms such as "ma'am" or "sir". Within LGBTQ spaces, such terms are not necessarily restricted to particular genders (e.g. gay men using *she/her*, butch women preferring "sir"); however, outside of such spaces they are intrinsically tied to gender rather than presentation. If one is called "ma'am" by a stranger, it is reasonable to assume one is being read as female, rather than as a feminine man or non-binary person. For this reason, the discussion of external perception below is focused less on specifically how one *is* gendered in public and more on the contexts in which speakers are *concerned* about being gendered a particular way (e.g. desiring to pass at work).

6.4.1 *Participant variation in passing and perception*

First session-responses to external gendered perception of one's voice were substantially more uniform than internal perception, and speakers produced fewer ambivalent responses. Overall, even speakers who themselves perceived their voices as *masculine*, *male*, or *androgynous*, or expressed ambiguity towards one or more of these, overwhelmingly reported that others did not. The sole initial atypical response to Q22 ("My voice passes as a cis male's") and Q23 ("My voice is perceived as male by others"), and only ambivalent response to Q26 ("...feminine...") was Oz, corresponding to greater initial time on HRT. Both Oz and Southern Boy responded atypically for initial Q24 ("...female...") and Q25 ("...masculine..."); Southern Boy

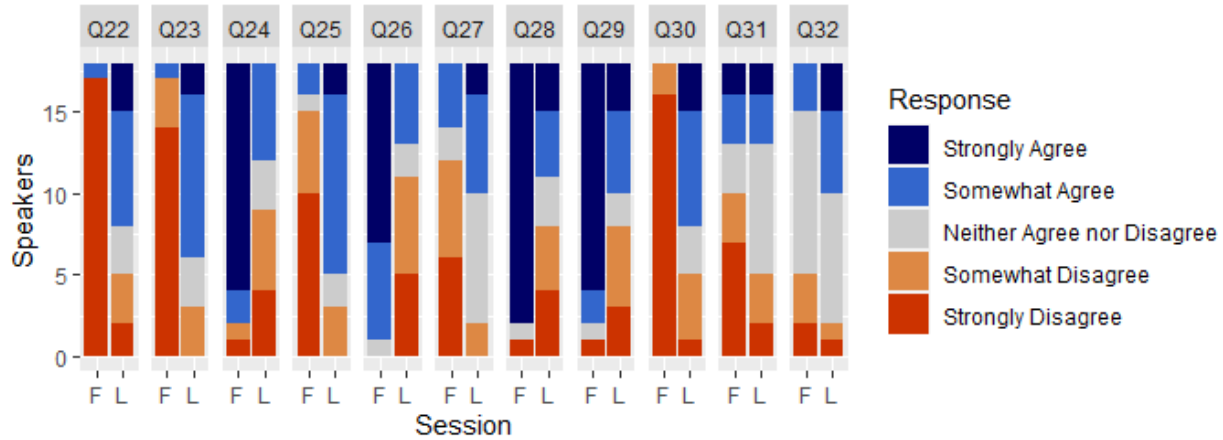


Figure 6.3: Distribution of (F)irst and (L)ast-session responses to all External Perception questions exhibiting significant cross-participant change over time:

- Q22. My voice passes as a cis male's.
- Q23. My voice is perceived as male by others.
- Q24. My voice is perceived as female by others.
- Q25. My voice is perceived as masculine by others.
- Q26. My voice is perceived as feminine by others.
- Q27. My voice is perceived as androgynous by others.
- Q28. I am frequently misgendered over the phone.
- Q29. I am frequently misgendered when speaking with others in person.
- Q30. People tend to gender me correctly after hearing my voice.
- Q31. People can tell I am trans based on my voice.
- Q32. People assume I am not straight based on my voice.

prior to the study already found others to read their voice as *masculine* and not as *female*, corresponding with agreement to Q14 (self-perception of one's voice as *male*) in Session 1.

By the final session, most agreed that their voice was perceived as both *male* and *masculine* by others, with only 3/18 (Kel, Marshmallow[†], Wizard) responding atypically to the former and 3/18 (Kevin, Marshmallow[†], Ty) to the latter. Again we may attribute Marshmallow's responses to fewer sessions; he and Wizard notably were the only ones of this group to have expressed general dissatisfaction (by way of lingering dysphoria). Kevin and Ty had also reported relatively low internally-perceived vocal *masculinity*, and high *femininity*, respectively, with both responding ambivalently to vocal maleness (Q14). In general, those showing lower internally-perceived vocal *maleness* or *masculinity* thus also appear more likely to experience others perceiving their voices as relatively less *male/masculine*.

However, even those whose voices were perceived as *male* by the last session did not necessarily report them passing as those of cisgender men. Many participants (8/18) did not agree with Q22 in the final session, with 5/18 (Arrow, Chimmy, Kevin, Lizard, and Southern Boy) explicitly expressing disagreement. Of these speakers, all reported their voices being perceived as *male* (Q23) (save for Southern Boy, who was ambivalent). Thus we find a large group whose voices are reported as *being perceived as male*, but not *passing as cisgender*. Of those who did report their voices passing as cisgender (ATM, Freddy, Gareth, Kevin, Marshmallow[†], Raine White, Renaud, Saint, Ty, and Wizard), we see variation in whether their voices are also specifically perceived as *masculine*, with Kevin, Marshmallow[†], and Ty in particular disagreeing with Q25; for these speakers, sounding *masculine* is not necessarily a prerequisite for passing vocally as cis.¹¹

Regarding perceived femaleness (Q24) and femininity (Q26), speakers during Session 1 near-universally reported that others read their voices as *female*; the exceptions here are Oz and Southern Boy, corresponding to that pair's higher vocal perception as *masculine* pre-HRT. By their final sessions, half of participants (9/18) conclusively reported that their voice was perceived as *female* by others; six disagreed (Arrow, Chimmy, Kel, Kevin, Marshmallow[†], Wizard) while a further three (Gareth, Renaud, Southern Boy) were ambivalent. With the exception of Gareth, who reported perceived *femaleness* but not *femininity*, this is a close overlap between final-session Q24 and Q26. Thus, when it comes to external perception, it appears that a *feminine* voice is also seen as a *female* one.

Worth noting, however, is that there is also some overlap with Q22; Kevin, Marshmallow[†], and Wizard all reported that their voices passed as a cis male's *and* that others found they sounded *female*. There is also overlap with the Q12 and Q15 responses regarding internal perception of femininity, though not complete. While most who reported that others read

11. What this means for each speaker differs depending on their personal definitions of "sounding masculine", but it is worth referencing again Southern Boy's discussion (S1|W3, §6.3.2) of desiring voice lowering to a pitch where they would be read as a man without feeling the need to change their inflectional patterns to a stereotypically masculine monotone.

	Ar.	AT.	Ch.	El.	Fr.	Ga.	Ke.	Kv.	Li.	Ma.	Oz.	Pa.	Ra.	Re.	Sa.	So.	Ty.	Wi.	Summary			
Q22											●								(F) Voice passes as cis male			
Q22	○		○○	x				○○	○		x	x					○		(L) V. does not pass as cis male			
Q23											●								(F) V. perceived as male			
Q23			x				○			○							x	x	○	(L) V. not perceived as male		
Q24											○						○○			(F) V. not perceived as female		
Q24	●		●			x	●	●		●					x				●	(L) V. perceived as female		
Q25	x										●								●	(F) V. perceived as masculine		
Q25			x					○		○									x	○	(L) V. not perceived as masc.	
Q26											x										(F) V. not perceived as feminine	
Q26	●		x				●	●		●									●		x	(L) V. perceived as feminine
Q27	●	●						●				x	●	x								(F) V. perceived as andr.
Q27		x			○	x			x		○	x	x	x	x	x						(L) V. not perceived as andr.

Table 6.8: Notable "External Satisfaction" Q22-27 responses by participant, in both (F)irst and (L)ast sessions. Bullets indicate response strength: ○○ *strongly disagree* | ○ *disagree* | x *neither agree nor disagree* | ● *agree* | ●● *strongly agree*.

their voices as *female/feminine* also perceived their own voices to be so, Chimmy did not, with Kevin and Wizard showing at most ambivalence to the latter. Thus, a speaker's own gendered interpretation of their voice is not necessarily the same as the one created by others.

Even among those who do express relatively high (internal) vocal femaleness/femininity, they may view *different aspects of their voice* as female/feminine compared to other people. For example, ATM responded typically in final-session external perception, reporting overall that people read his voice as both male and masculine and that it typically passed for that of a cisgender man; he found, however, that his laugh specifically appeared to change how others gendered him. Though giving no indicators of general dissatisfaction, ATM described relatively high internally perceived vocal femininity in his final session; his laugh, however, was not a feature he believed contributed to this (ATM S12|W57).

I have had multiple instances where I pass without trying. And then all of a sudden, they're like, "Oh!" - I'll laugh, it's my laugh, usually. And they're like, "Oh sorry, was that [name] or [feminine variant of name]". And I'm like, I quit [*laugh*]. I don't think my laugh sounds very feminine!

(ATM S12|W57)

We see differences not only between how speakers and their interlocutors or strangers perceive gender in their voice, but in which aspects of that voice those individuals are using

to inform those perceptions. Whether relatively high post-HRT rates of being perceived by others as vocally *female* or *feminine* are necessarily indicative of dissatisfaction depends, like internal perception, on individual speakers' goals and identity. As said, several speakers explicitly said that they did not wish to pass as cisgender men (see §3.4.2), while others did, with the latter group including both those for whom a cisgender male voice best-reflects how they identify (e.g. Raine White) as well as those who do not identify as male but still desire to pass as such for safety or convenience (e.g. Wizard S1|W1).

I'm really just hoping to feel more comfortable in my own skin, er, I guess in my own vocal chords. [*laugh*] I don't necessarily want people to hear my voice and think, "Oh, that's a man." [...] [But] for strangers, that is honestly kind of the goal, like if I'm going in a grocery store and interacting with people, I don't want to have to say "Well, I'm genderqueer non-binary and this is my-", like, I'd rather people perceive me as a man, just to get in and get out. (Wizard S1|W1)

Note the contrast here with Lizard's commentary about initially trying to pass as female for the same reasons (§6.3.2 Lizard S1|W4). This is further complicated by perceived vocal androgyny, which is an explicit goal for some speakers (e.g. Arrow) while for others it is a transient state before being read as male. Not all non-binary speakers desire an androgynous presentation; they may view themselves as *feminine*, *masculine*, neither, both, or something else entirely. For those who do desire to be perceived androgynously, however, since society largely frames non-binary identity and gender-neutrality both as non-normative, they are also therefore viewed as things *to be transitioned to*. Thus, for such speakers, being perceived specifically as *non-binary* is concurrent with being read as LBTGQ (either as trans or incorrectly as a butch lesbian, etc.); there is no such thing as "passing" as non-binary.

There is full awareness among non-binary speakers that interlocutors will not typically assume non-binary identity. Additionally, as suggested above, the *female-versus-feminine* distinction may not be one made in the voice by listeners when interacting with a speaker

who, being within the first year of HRT, may still be consistently read as *female*. Thus, even among non-binary speakers who would describe themselves as *feminine*, we see a general preference that others assume them to be the other *binary* gender, rather than assuming them to be the gender they were assigned at birth (Elias S1|W0).

If I had to choose between, like, being confusing, and being masc or being femme, I'm honestly pretty equally good with being confusing or masculine, and just, definitely don't wanna be perceived as femme. Um, I mean, I *am* femme [...] but I know, especially the people who aren't, like, knowledgeable about gender stuff, they're just going to go one way or the other. So I would rather those people perceive me as male.

(Elias S1|W0)

As for perceived androgyny, in Q27 ("My voice is perceived as androgynous by others") we see the same trend towards more positive responses as with maleness/masculinity. Arrow, ATM, Kevin, and Raine White reported relatively high perceived vocal androgyny in Session 1; of these, only Raine White identified his *own* voice as androgynous. Thus in androgyny, too, we see a difference in external versus internal perception. In the final session, most either claimed their voice was viewed as androgynous (8/10) or were ambivalent (8/10). I suggest this high ambivalence for Q27 to be due to the same reason behind low ambivalence in Q22-26: it is difficult to know when others are reading one's voice as androgynous. Knowing how one is gendered by others relies on input (e.g. pronouns, honorific terms of address, etc.); one can be quite certain of being read as *male* or *female*. Rarely does that input confirm androgyny, as strangers are unlikely to admit they aren't certain how to gender a speaker.

One indicator of perceived androgyny is how others use one's name. ATM, for example, has a male-typical first name with a common female-typical variant and a gender-neutral shortened form. He saw a shift over time on HRT: initially, strangers would mishear his introduction and assume the feminine variant; later, others would prefer the gender-neutral form, though accompanied by continued misgendering as female (ATM S3|W11).

They're like, oh, you're going for a gender-neutral name. And that's absolutely *not* what I'm doing. [...] It's like, oh, do I have permission to misgender you because I'm neutering your name? (ATM S3|W11)

The mishearing or misinterpretation of his name was a recurrent issue for ATM (c.f. S12|W57 above, S1|W2 in §6.4.2), with the distress of this leading eventually to him disallowing others who were not close friends from using the shortened form, but he was not alone. Numerous other speakers described instances where others recognized *masculinity*, *maleness*, or *androgyny* in their voice or presentation and assumed them to be gender-non-conforming women — or, more rarely, non-binary (when the speaker was not).¹² Thus, to those for whom it is not a goal, assumed gender neutrality can itself be form of misgendering. Similarly, several participants reported that they were more often asked their preferred pronouns over the course of transition (i.e. as their voice and presentation began to cue interlocutors not to assume the use of *she/her*). Though asking for one's pronouns is relatively common in many modern LGBTQ spaces, a few speakers found that they were the only person in a group who was asked; this was welcomed by some, but met with frustration by others who wished to be read unambiguously as male and who saw being the only one asked as indicative of unwanted perceived androgyny.

There were some instances in which others explicitly mentioned perceiving a participant's voice ambiguously or androgynously (Chimmy S12|W71). Outside of these isolated situations, however, determining that one's voice is read as androgynous may come from mixed instances (being gendered differently by different people) or from experiences of strangers noticeably second-guessing forms of address or expressing uncertainty towards a speaker's gender.

The closest I've gotten to actual feedback on my voice was [...] somebody who's told me that my voice doesn't sound either male or, or female. Um, and they gave me that

12. While ATM was "highly likely" to use the term *non-binary* by Session 12, in Session 3 he was ambivalent, with the shift from "(highly) unlikely" to "(highly) likely" occurring between Sessions 3-5.

comment knowing that I do get *ma'am*'ed on the phone and don't like it. So yeah, it - [*laugh*] - it was interesting to get that feedback. It wasn't like, "Oh no, of course you sound like a man, don't worry about it." It was "No, you're in-between."

(Chimmy S12|W71)

Several participants describe learning to recognize this uncertainty and use it to determine how they are being read, combining external feedback with internal physical sensation to gauge how their speaking voice comes across to others. For example, Marshmallow by his sixth session was very rarely perceived as male, and was also experiencing some physical vocal strain; he used both of these elements to fine-tune his voice (Marshmallow S6|W27).

I can catch other people's responses. There's a different way that people interact with a man, versus the way they interact with a woman - there's different words, phrases, tones, and I'm learning to sense the difference in their responses, and that tells me how I'm coming across. So it's between [that and] feeling - feeling the strain, especially in my vocal chords and - and throat, losing the vibrational feedback in the chest, um, those two things. [...] And then that feedback, that response from them, kinda all gears me towards what I'm achieving, or not accomplishing well, in that moment. Since I don't have the - the memory yet of what it should sound like, I'm re-tuning and re-learning that.

(Marshmallow S6|W27)

In this way, external perception becomes a cue not only for how one is being gendered, but for overall vocal health. Similarly, some participants reported changes in others' perceptions of their voice related to the effects of HRT, but not directly to gender. The most common of these was loudness, with several reporting that others had a difficult time hearing or understanding them in noisy spaces (see Chimmy's concerns about volume, §3.4.2), or that they had difficulty projecting their voice following some time on HRT.

6.4.2 Participant experiences with misgendering and assumed queerness

Misgendering takes many forms, from the minute to the egregious. It may be briefly defined as treatment at odds with one's gender identity, ranging from a stranger in passing using incorrect pronouns to systematic exclusion from gendered spaces. Instances of misgendering may be intentional, as in an unsupportive environment where the actor is aware of the speaker's gender identity but refuses to acknowledge it. It may also be entirely unintentional, where the actor simply does not know the speaker's identity and, upon a first interaction, guesses incorrectly. It is this later case of misgendering to which this questionnaire largely refers, as this is the case that depends on how trans individuals are perceived by others.¹³

Though participants reported voice as a major component in whether they are misgendered, voice is far from the only relevant factor; initial perception of gender is based heavily on physical appearance. While testosterone does have various physical effects aside from those on the voice (e.g. growth of facial hair, redistribution of body fat), many of these take place on a time scale larger than the first year. Presentation, attire, and behavior are further factors; as such, we must not assume that continued in-person misgendering reported at the end of the first year on HRT is exclusively due to speakers' voices. Instead, participants found that when their appearance was ambiguous, their voice served as a second gender cue for listeners. Early in the study, this resulted in several participants occasionally being assumed *male* in conversation, only to have their interlocutor proceed to misgender them upon hearing their voice (typically following an apology, as the interlocutor believed they had in fact been *initially* misgendering a *female* speaker). Near the end of the study, some found the opposite had begun to occur, where they were read as female until they spoke; this is the scenario referenced in Q30. Still others found that their changed voice had little effect, as they were still perceived as female regardless, based on other physical characteristics.

13. Actors engaging in intentional misgendering do not do so because of different initial perceptions, but because of a belief that the individual is *not* the gender they are; a trans man may largely pass as cis but still experience intentional misgendering from those who know of, and take issue with, him being trans.

	<i>Ar.</i>	<i>AT.</i>	<i>Ch.</i>	<i>El.</i>	<i>Fr.</i>	<i>Ga.</i>	<i>Ke.</i>	<i>Kv.</i>	<i>Li.</i>	<i>Ma.</i>	<i>Oz.</i>	<i>Pa.</i>	<i>Ra.</i>	<i>Re.</i>	<i>Sa.</i>	<i>So.</i>	<i>Ty.</i>	<i>Wi.</i>	Summary	
Q28						○						x							(F) Not misgendered on phone	
Q28	x		●●	x			●	●●		●	●●	x				●	●		(L) Misgendered on phone	
Q29	x					○○													(F) Not misgendered in person	
Q29			●●		●	x	●●	●		●	●					●	○	●●	(L) Misgendered in person	
Q30	x		○				○	○		○	○○					x	x		(L) Misgendered by voice	
Q31	x		x	●		x	●●	●	x	●	x	x			x	x			●●	(L) Assumed trans
Q32	x	●	x	●	x	x	●●	●	x	●●		●	x	x	x	●			●●	(L) Assumed not straight

Table 6.9: Notable "External Satisfaction" Q28-32 responses by participant, in both (F)irst and (L)ast sessions. Bullets indicate response strength: ○○ *strongly disagree* | ○ *disagree* | x *neither agree nor disagree* | ● *agree* | ●● *strongly agree*.

Nearly all participants reported frequent in-person misgendering (Q29) in-person prior to the study, with Gareth as the only atypical responder and Arrow as ambivalent. Gareth's disagreement may be attributed to their relatively minimal social interaction, wherein there were few circumstances in which they could be misgendered by others. The ambivalence shown by Arrow is a pattern continuing throughout these questions, which prompts a question to which we have returned several times: how does the concept of "misgendering" interact with *non-binary* experiences, in a world built upon an assumed *male-female* binary of gender? In Oz's words, "I kind of don't expect anybody to gender me correctly on first sight, because [...] [being] non-binary is kind of complicated with misgendering right now" (Oz S1|W14). Many non-binary participants did not anticipate ever being correctly gendered outside of LGBTQ spaces or places where awareness of non-binary identities is common and expected. It is partially for this reason that we find non-binary speakers desiring to pass as male, for the sake of convenience or safety (see Wizard S1|W1 above). Ultimately, what a speaker considers misgendering is largely personal, particularly for those with identities that are not wholly male.

Overall, while we did see the expected significant reduction in in-person misgendering, a total of 8/18 participants (Chimmy, Freddy, Kel, Kevin, Marshmallow[†], Oz, Southern Boy, and Wizard) responded atypically to Q29 in the final session. This group includes many who reported lower general satisfaction (Q1, Q10), as well as high self-perception of vocal

femaleness in their last session (Q15). Of this group, Chimmy, Kevin, and Southern Boy also reported their voice passing as a cis man's (Q22), with only Kel, Marshmallow[†], and Wizard reporting it *not* being perceived as *male* in general (Q23); thus, among those who are still misgendered after the final session are many whose voices *otherwise are read as male*.

Initially, this not unexpected, given the importance of non-voice factors in passing. To isolate voice, we may look at Q28 ("I am frequently misgendered over the phone") and Q30 ("People tend to gender me correctly after hearing my voice"). Doing so, however, reveals three points: first, that 7/18 participants *still report being misgendered over the phone* after one year of HRT; second, that 5/18 report others not gendering them correctly after hearing their voice; third, that the latter group is a subset of the former, and both (nearly) a subset of those who are still misgendered in person. As with (Q29), every speaker reported initial misgendering over the phone (Q28) with the exception of Gareth (and Panda showing ambivalence). This consists of the same use of incorrect pronouns or terms of address as in-person misgendering, with several participants specifically mentioned issues with the mishearing of their names; ATM, who has a highly masculine name, describes what he calls a "pretty typical experience" (ATM S1|W2).

I'd said, "Hey, I'm calling to make an appointment." She goes "Okay, who's it for?" I said, "For [name]," and she goes, "Oh, for someone else." And I was trying my hardest to make my voice sound like, [at a lower pitch] "I'm [name] and I'm here to get an [...] appointment." [laugh] And, uh, at the end of the call she goes, "So, I just need to verify that the spelling of your name is, uh, is it [name] or [typically feminine variant of name]." So I was like [groan], and I say "Oh, I'm a trans man." And she says, "Oh, thank you for telling me that." And I was like, "You're welcome. I didn't want to have to out myself on a phone call to someone. But you're welcome." (ATM S1|W2)

Except for Ty, who indicated misgendering over the phone but not in person, all those still frequently experiencing the former also experienced the latter. Freddy and Wizard alone

reported misgendering in person but not on the phone; this is somewhat surprising, as both also were among those who did not believe their voices passed as a cis male at the end of the study, and Wizard further did not think his was perceived as male at all. As part of his work, Freddy spends a substantial part of the day in phone conversations with strangers, and described throughout the study a very slow shift in how those individuals referred to him. In the first session, no longer being misgendered specifically on the phone was a major priority, as it was a near-constant occurrence (Freddy S1|S1).

My job is answering the phone all day long. And I am misgendered all day long as a result. [...] I'll say, hello, my name is [*name*], how can I help you, and people will get very hung up on how I have a masculine name but a feminine voice. And I get pushback on my name constantly, people are like "Oh, is it short for [*typically feminine variant of name*]?" I'm like, no, and why do you care? So that's been a huge source of dysphoria and discomfort. [...] So I'm just stuck, being perceived by all the customers, it's just stuck, it's like I'm playing a role. (Freddy S1|W1)

After approximately 6 months on HRT, Freddy described this experience worsening; his voice had slowly begun to be gendered correctly over the phone, and so when it was not, those instances stood out and *felt* worse in comparison. By the final session, he was gendered correctly on the phone "99% of the time". Chimmy, in contrast, saw very little improvement in being gendered correctly over the phone, which was a pervasive source of stress throughout the year. Ultimately, HRT alone was insufficient in addressing this, which only began to see some relief in the final session after having begun voice therapy (Chimmy S12|W71).

I feel like I'm – I'm not *ma'am*'d 100% of the time on the phone anymore. [...] That was a big, big deal. [...] I think it is the result of the vocal therapy, um, I'm not getting *sir*'d on the phone or anything, but it's just not there as consistently anymore, which is a huge relief. [...] I don't know if I can expect my voice to change that much more, um, but I think I can practice using my voice in – in ways that are less at odds with my gender. I feel like it's mostly on me at this point. It's practice. (Chimmy S12|W71)

The final component to external perception is whether speakers find that others specifically *recognize them as trans* from their voice (Q31), and/or *assume them to not be straight* (Q32). These questions stood out to speakers; while many were aware of the phenomenon of transmasculine voices being read as gay (and indicated having experienced it) several others remarked that they hadn't considered the possibility before ("That question brings me joy," -Wizard, S11|W66), and some who had expressed concerns about femininity in their voice commented that they would not mind if they were perceived as feminine, gay men.

Perception as both *trans* and *not straight* was common; only five speakers (ATM, Freddy, Raine White, Renaud, and Ty) disagreed with Q31 in their last session, and only two (Oz, Ty) with Q32. The linear regression did not show significant correlation with time for either question, while the mixed effects model showed a significant but extremely tiny effect of time for both. The questions themselves prompted substantial discussion with participants on the topics of passing and what "being read as trans" actually means, differing depending on situation and presentation. One situation described was that of a pre-transition transmasculine person who, particularly in LGBTQ spaces where they are comfortable expressing themselves genuinely, finds that others assume or suspect they are trans despite not disclosing.

Another situation was that of someone who, after some time on HRT, has a voice typically read as *male*, but who does not pass physically (either due to inability or to not desiring to). They may initially be read as female in public, only to be met with confusion when they speak, as their voice is perceived as incongruous with their appearance. One notable phenomenon in this case is that of others identifying them as non-passing trans men, but as gender-non-conforming women or even as trans women. Oz in Session 9 described an instance of the latter at the DMV; they are non-binary and do not consider themselves a man, though they often pass as male in public, but their licence at the time still read "female". This, in combination with their voice, lead to interactions with a driving instructor that were, thankfully, only awkward rather than dangerous.

The final situation described is that of a speaker who passes as a cisgender man, but whose voice leads others to believe they are trans; we may relate this to the distinction between believing one's voice is perceived as *male* and believing it specifically passes as a *cis male voice*. This was mentioned particularly by speakers who observed a "trans accent" (see §2.2.3), such as Oz and Kel. This latter situation was less common; more often, rather than "sounding trans", participants (particularly those who described others perceiving their voices as *feminine*, see Q26) reported "sounding gay", in which self-identified feminine qualities of their voice (again inflectional patterns were given particular import here) led others to assume they were not straight. As discussed in §3.4.2, the majority of speakers (15/18) did in fact use the term *gay* to refer to themselves; several were married to men or had male partners, and openly embraced being perceived as gay men. Notably, while there is an enormous variety of orientations outside of only gay or straight, participants' elaborations on Q32 responses typically only referenced perception as gay; they never found, for example, that strangers heard their voices and immediately assumed they were bisexual.

It is difficult to determine the role that voice specifically played in this perception, until an accompanying perceptual study may be performed. It was rare for participants to respond conclusively to either Q31 or Q32, with high rates of ambivalence (8/18) for both. This is likely due to some of the same mechanisms behind high ambivalence in the androgyny-related questions: it can be difficult for a trans person to know for certain whether another person assumes they are trans, or not straight. Depending on environment, this knowledge may be crucial, such as in areas and situations where being recognized as such puts one at risk of confrontation or violence. Most speakers were fortunate enough not to live in such an environment, though several did choose to go stealth in their workplaces or colleges.

Overall, participants after the first year of HRT showed high rates of continued misgendering both in-person, which may be informed by non-vocal characteristics, and over the phone, which may not. In fact, all those in their final session who agreed with Q28 (mis-

gendering over the phone) also agreed with Q29 (misgendering in-person), save for Ty, who showed ambivalence to the latter. This suggests that it is not the case that this continued misgendering is wholly the result of other physical characteristics; else we would expect far more to be misgendered in-person than over the phone. Instead, these speakers' voices *on their own* do not lead to an accurate assumption of gender from others; they have, in effect, not reached Step 4 of the "expected transmasculine vocal narrative". While for some this may not a source of distress, or even be expected due to the difficulties in being assumed non-binary (e.g. Oz), for others it may be a continued source of vocal dissatisfaction (e.g. Chimmy); we must take into account speakers individual desires on the matter.

6.5 Results: Desire for Future Change

This final section, collapses the desired internal and external perceptions of one's voice; i.e. there is a single Q33 "I wish my voice sounded more masculine" rather than separate questions asking "...more masculine *to myself*" versus "...*to others*". This was done both to optimize survey length and because while making this distinction in perception is useful to understanding participant motivations and satisfaction, ultimately any vocal strategies implemented to alter how the voice sounds in one respect is going to affect the other.

Significant change was seen in participants' desires for their voice to sound *masculine*, *male*, *androgynous*, and *feminine* (Q33-36), with all but the latter showing an inverse correlation with time on HRT (Table 6.10). There was no significant change in the response to Q37 "I wish my voice sounded more female". Similarly, a negative correlation was seen in the desire for one's voice to (continue to) change significantly (Q38), as expected. There was no overall pattern of change in the desire for vocal therapy, vocal surgery, for one's voice to *pass as a cis man's*, or for one's voice to be *read as trans* (Q39-42); unexpectedly, there was a slight negative correlation with the desire for one's voice to be read as *non-straight* (Q43). The very slight positive correlation between time on HRT and Q34 "I wish my voice

#	Desire for Further Change	r	β
33	I wish my voice sounded more masculine.	-0.406**	-0.171**
34	I wish my voice sounded more feminine.	0.139*	0.003*
35	I wish my voice sounded more androgynous.	-0.160*	-0.015**
36	I wish my voice sounded more male.	-0.378**	-0.016**
37	I wish my voice sounded more female.	0.060	0.002
38	I do not want my voice to change significantly.	0.437**	0.024**
39	I am considering vocal therapy to change how my voice sounds.	0.104	-0.001
40	I am considering vocal surgery to change how my voice sounds.	0.007	-0.001
41	I want my voice to pass as a cis male's.	-0.079	-0.001
42	I want people to be able to tell I am trans based on my voice.	-0.103	-0.004
43	I want people to assume I am not straight based on my voice.	-0.153*	-0.011**

Table 6.10: Correlation coefficients between "Desire for Change" question results and time (r) and fixed effects estimate of time (β) all participants; higher r indicates stronger correlation, positive/negative r and β indicate direction of correlation. Significant effects are in bold; stars indicate level of significance (* $p \leq 0.05$, ** $p \leq 0.001$).

sounded more feminine" is notable, particularly considering that no such correlation was seen for sounding more *female*. This again supports the importance of nuance in wording and in recognizing how both presentation and gender identity are recognized in the voice; as discussed in §3.4.2, numerous participants by the final session (specifically Arrow, Gareth, Lizard, Panda, South, Ty, and Wizard) did use the term *feminine* to describe themselves, regardless of specific gender identity.

6.5.1 Participant variation and desire for future change

Despite many questions not showing change over time, they provide valuable context for prior responses regarding both internal and external perception. Whether one *desires* their voice to sound more masculine (Q33), for example, can recontextualize a previous observation that their voice does not sound masculine to others (Q25); thus they are still worth consideration.

Participant desire for gendered change

Just as with perception, in desire for future change we find a distinction between Q33 ("I wish my voice sounded more masculine") and Q36 ("...more male"). In their initial session,

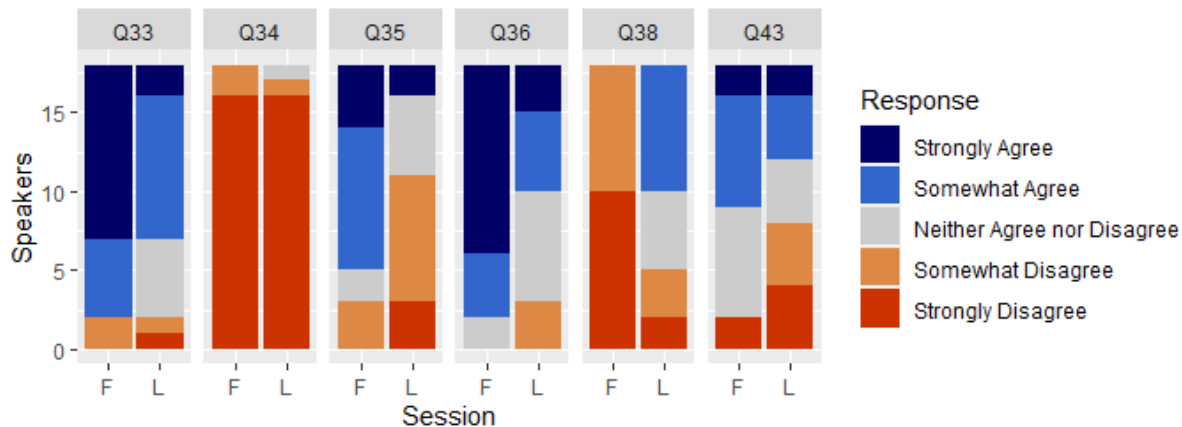


Figure 6.4: Distribution of (F)irst and (L)ast-session responses to all Desire for Change questions exhibiting significant cross-participant change over time:

- Q33. I wish my voice sounded more masculine.
- Q34. I wish my voice sounded more feminine.
- Q35. I wish my voice sounded more androgynous.
- Q36. I wish my voice sounded more male.
- Q38. I do not want my voice to change significantly.
- Q43. I want people to assume I am not straight based on my voice.

every participant save for Arrow and Oz responded in agreement for both questions, with this pair both disagreeing to the former and showing ambivalence to the latter. Oz was already largely content with how their voice had masculinized (see §6.3.1), while Arrow explicitly expressed initial desire for an androgynous voice. This was related in part to not thinking of their voice in gendered terms (concurrent with a non-binary identity), and in part to the avoidance of certain vocal characteristics (see Arrow S1|W8, §2.2.2). This sentiment was shared by other participants who did respond affirmatively to initial Q33, but who described not wanting to have particular male-typical vocal styles (e.g. "sounding like a dudebro").

By the final session, we find that many participants still desire, or would be ambivalent to, further vocal change. 11/18 speakers agreed with Q33 after a year of HRT, indicating that they still wished to sound more *masculine*, with a further 5/18 ambivalent; only Freddy and Lizard disagreed. Notably, while both Freddy and Lizard are speakers who have shown overall high post-study vocal satisfaction (responding typically throughout the final-session General Satisfaction block), others with high general satisfaction (e.g. Kevin, Saint, Panda,

	Ar.	AT.	Ch.	El.	Fr.	Ga.	Ke.	Kv.	Li.	Ma.	Oz.	Pa.	Ra.	Re.	Sa.	So.	Ty.	Wi.	Summary	
Q33					o	x			oo		x	x	x	x					(L) Low desire to sound more masc.	
Q36				x	o	x			o		x	o	x	x	x	x			(L) Low d. to sound more male	
Q35	••		x					x			x	x					x	••	(L) High d. to sound more andr.	
Q38	•	•	•	x		x	••	•		••	x		x	x	x	x		•	••	(L) Continued d. for vocal change
Q39	•		••			••	••	•			•			•	x			•	•	(L) Considering vocal therapy
Q41	o							x	o		x	x								(L) Low d. to pass as cis
Q42	x		x	•						•			x	x	x	x			x	(L) D. to be assumed trans
Q43	x	••		•	••	x	••	•	x	••		x	••	••	•	•			••	(L) D. to be assumed not straight

Table 6.11: Notable "Future Change" Q33-43 responses by participant, in both (F)irst and (L)ast sessions. Bullets indicate response strength: oo *strongly disagree* | o *disagree* | x *neither agree nor disagree* | • *agree* | •• *strongly agree*.

Raine White) still responded positively or ambiguously to Q33. This includes many who did report that their voice sounded *male* or *masculine* to themselves, or to others.

This highlights an important consideration in discussion of changing transmasculine voices; just as seen with the lingering dysphoria reported in §6.2.1 even among those who say they are happy with their voices, satisfaction (as with *maleness*, *masculinity*, etc.) is not all-or-nothing. Of the 11/18 participants who still desired to sound more *masculine* post-study, 8 said that they perceived their voices as *masculine* at that time, and 6 that others did so. For these speakers, their voices *were heard as masculine*; we cannot describe any low satisfaction among this group as the result of failure to achieve vocal masculinity. They still desired, however, for their voices to become more so. Some expressed the hope that they would see further biologically-driven vocal change from HRT, but all acknowledged that the majority of such change they could expect had already occurred, with most describing a future focus on voice training efforts and behaviorally-driven vocal change.

The final-session pattern for Q36 ("I wish my voice sounded more male") is much more similar to Q33 ("I wish my voice sounded more masculine") than their initial sessions, with substantial overlap in responses and only Freddy, Lizard, and Panda disagreeing. Just as with Q11/Q14, speakers appear overall more ambivalent when framing perceived vocal gender as *male* rather than *masculine*. We may note the distinction here of Panda desiring to sound more *masculine*, but not necessarily more *male*.

While there is no significant change over time for Q41 ("I want my voice to pass as a cis male's"), as participant responses remained generally positive in both initial (13/18) and final (12/18) sessions, it is still worth discussion. All three of Q33, Q36, and Q41 are most relevant in the context of the external perception question set (primarily Q25, Q23, and Q22, respectively) as they contextualize whether a negative response to these questions may be associated with lower satisfaction. Q33 and Q36 collapse desire for greater internal/external perception as more vocally *masculine/male*, but by referencing passing, Q41 explicitly pairs with Q22, and so the two may be directly compared.

Of the 12 speakers who by their final session desired their voices to pass as those of cisgender men (ATM, Chimmy, Elias, Freddy, Gareth, Kel, Marshmallow, Raine White, Renaud, Southern Boy, Ty, Wizard), two (Chimmy and Southern Boy) reported that this was not the case (via a negative Q22 response) while Elias was ambivalent. Of those ambivalent towards passing (Kevin, Oz, Panda, Saint), Kevin responded negatively to Q22, while Oz and Panda were ambivalent. Only two speakers, Arrow and Lizard, explicitly said in their final sessions that they did not desire to pass vocally as cisgender; though Lizard is showed high general satisfaction (responding typically throughout Q1-9) alongside nearly entirely-typical responses in external perception (i.e. his voice were generally perceived as *male* and *masculine*, neither *female* nor *feminine*, and he was typically gendered correctly both in-person and on the phone), Arrow was more ambivalent in several of these, additionally reporting lingering vocal dysphoria.

Q35 ("I wish my voice sounded more androgynous") saw significant change over the course of the study; in Session 1, 14/18 are in agreement, excepting Freddy, Kevin, Lizard, and the ambivalent Southern Boy. As discussed previously regarding vocal androgyny, despite its frequent framing as a midpoint between *male/female* or *masculine/feminine* perception, there are numerous reasons why it may not be desired by a speaker currently frequently read as *female*. Both Freddy and Kevin, for example, notably expressed a more binary identity

in Session 1 (though both would be "somewhat likely" to call themselves non-binary by the end of the study), with Kevin then explicitly citing a desire to pass as cis (see Kevin S1|W2, §2.1.5) and Freddy describing distress resulting from assumed androgyny by others.

By their last session, 11/18 participants did not desire to sound more androgynous. Only Arrow and Wizard responded atypically, both "strongly agreeing" with the sentiment, while a further 5/18 (Chimmy, Kevin, Oz, Panda, Southern Boy) were ambivalent. Depending on speaker, this change may be either because they had already reached a desired vocal androgyny (e.g. Elias), or because they had reached or would prefer to reach greater perceived vocal *maleness* or *masculinity*, such that *androgyny* now was undesired. Ty, for example, shifted from strong disagreement with all of Q11, Q13, and Q14 to ambivalence towards the former two and agreement with the latter; i.e. his voice grew in his own perception more *male*, *masculine*, and *androgynous*. However, while he also showed matching initial agreement with all three of Q33, Q34, and Q35, by the last session he responded negatively to Q34. In his own, more qualitative words: "I think I'm pretty much in an androgynous zone, which I don't totally mind. But it wouldn't hurt if it was a little lower." (Ty S11|W63)

We see no change in Q37 ("I wish my voice sounded more female") over time. The only non-negative response to this across the study was Southern Boy's ambivalence in Session 12. However, Q34 ("I wish my voice sounded more feminine") did show an extremely slight but significant *increase* over time. Examining only first and last sessions across participants does not make this change apparent; like Q37, the only non-negative response in those sessions was that of Southern Boy, and the majority chose "strongly disagree" in both sessions. Looking at overall month-to-month responses, we do see that many participants who "strongly disagreed" for Q34 would come to "somewhat disagree" for one or more sessions in the second half of the study. We have not largely drawn a distinction in analysis between "somewhat" and "strongly" disagree; despite this, and despite these speakers ultimately returning to "strong" disagreement, this slight shift up resulted in a significant effect.

The effect is too small, and appears dependent on the responses of two few participants, to conclusively determine a pattern in *why* speakers wished their voices to sound more feminine. In the case of Southern Boy, it is worth noting that over the course of the study he had grown increasingly involved in his local drag scene, developing a drag queen persona including a stylized voice. While his drag persona's style of speech does not reflect his own preferred speech characteristics, it is notable that the only speaker to indicate a desire for greater vocal femininity is one actively involved in a *performance* of femininity, divorced from oneself and part of a character (Southern Boy S8|W44).

I was terrified at first that I was gonna have all kinds of dysphoria, especially with people calling me *she/her*, but like, something about putting everything on makes me a completely different person, and I don't mind any of that stuff. [...] I'm excited, because it gives me a chance to re-explore the things I used to enjoy, without jeopardizing my own sanity and dysphoria. (Southern Boy S8|W44)

Participant desire for outside intervention

There was no overall significant change in Q39 ("I am considering vocal therapy to change how my voice sounds") or Q40 ("I am considering vocal surgery to change how my voice sounds"); we may attribute this in part to wide participant variation. For Q39, participants fell into various groups: those who agreed throughout the study (Chimmy, Gareth, Kevin, Oz, Renaud, Ty, Wizard), those who disagreed throughout (Freddy, Lizard, Marshmallow, Panda), those who moved from disagreement to agreement (Arrow, Kel), and those who moved from agreement/ambivalence to disagreement (ATM, Elias, Raine White, Saint, Southern Boy). In the latter group, we find mostly those (save for Southern Boy) who showed high general satisfaction with their voices, though all but Raine White still expressed some desire to sound more male or masculine. In contrast, Arrow and Kel reported that their voices were still read as female by the end of the study (which none of the prior group did).

Not all those who wanted their voices to be differently-gendered post-study desired vocal therapy; gender-affirming voice therapy for transmasculine speakers is still relatively new, since as said, prior focus has been on transfeminine speakers who do not have the assistance of HRT's effects on the voice. Thus, impediments such as cost and availability play as much a role here as speakers personal desires for formal therapy. Chimmy is notable in having found therapy to address issues with being read as female over the phone (see Chimmy S12|W71, §6.4.2) that HRT alone could not. Rather, desire for vocal therapy more closely overlaps overall desire for continued vocal change (Q38); the speakers who do not feel they are entirely happy with the results of HRT are largely also those interested in therapy.

The desire for vocal surgery (Q40) was universally disagreed with in Session 1, with all but Renaud continuing to do so in their final session. Several expressed surprise that such surgeries existed; as discussed in §2.2.4, they are extremely rare, predominantly used for whom HRT is ineffective (due to medical complications) or not sufficiently so. This question was included because, while exploring online trans community spaces, I found discussion of such surgeries as a potential solution for those who desired lower voices, but not the other effects of HRT. While most participants desired the entire range of physical changes, it was not uncommon for voice lowering to be the primary reason they had begun, and some did indeed comment on having considered vocal surgery as an alternative (Saint S1|W1).

I'd say voice is one of the most important things to me, because when I first realized I was trans, I didn't really care about starting testosterone. [...] I was like, I'm just gonna get surgery to, to fix my voice, cause I don't wanna do – cause I don't think I want everything that testosterone offers, you know. So the voice was one of the main factors to why I actually started. (Saint S1|W2)

However, even those participants expressing some dissatisfaction with their voice did not tend to consider vocal surgery as an option. Renaud is the only exception, showing ambivalence to Q40; this follows a consistent pattern of ambivalent responses from Renaud.

Participant desire to be read as trans or not straight

Q42 ("I want people to be able to tell I am trans based on my voice") showed no significant change over time, though Q43 ("I want people assume I am not straight based on my voice") showed a significant decrease. As with their corresponding external satisfaction questions Q31 and Q32 (§6.4.2), these questions prompted discussion with participants regarding what it means to be assumed trans or not straight, and how they felt voice contributed to this.

In the first session, 8/18 participants reported not desiring to be read as trans based on their voices; a further 8/18 were ambivalent, with only two (Saint and Oz) responding atypically. Examination of month-to-month responses shows most participants with high variability in this particular question; we can expect a variety of factors any given month, including instances of misgendering or being correctly gendered, positive or negative reactions from strangers who read one as trans, change on HRT, and experiences within trans spaces to affect the extent to which one desires to be outwardly identifiable as trans. Very few speakers had strongly positive or negative responses over the course of the study, varying between ambivalence and agreement (Arrow, Gareth, Saint), ambivalence and disagreement (ATM, Kel), or both (Elias, Kevin, Marshmallow, Panda, Raine White, Wizard), while others remained ambivalent throughout (Chimmy, Renaud, Southern Boy).

By the last session, we again see two atypical responses to Q42, though this time from Elias and Marshmallow; Saint now was ambivalent, while Oz disagreed. Neither atypical response is particularly notable given the high month-to-month variation seen in both speakers, but Elias' discussion of the topic illustrates one possible source of this variation: a desire to be read as trans only situationally (Elias S12|W55).

I have a desire to definitely, like, pass more as cis and gender-conforming and stuff, in hetero or strange places. [...] My opposite's very drastically in the other direction, like, I wanna be visibly trans. But I'm not quite where I'm comfortable with that all the time, more just in queer-centered spaces. (Elias S12|W55)

As for Q43, 9/18 participants agreed in the initial session that they desired others to assume they were not straight based on their voices, with a further 7 ambivalent; the only atypical responders here were Marshmallow and Ty, both of whom strongly disagreed. Notably, of all participants, only Marshmallow identified as *straight* in this first session, with Ty instead preferring *gay* and *queer* among other terms. Thus, Marshmallow's preference not to be assumed not-straight is likely the result of being, in fact, straight, while Ty's is not. Ty, however, was also among those who consistently desired others not to recognize him as trans by his voice (Ty S8|W44). This quote came directly before the Ty S8|W44 quote in §2.2.3, as Ty goes on to hypothesize that the person who clocked him had picked up on some transmasculine-specific vocal style.

I realized more people are clocking me as trans. [...] Like, recently [someone at work] looked at me, and she goes, "Are you trans?" and I'm just like, "Yes." [...] She goes, "Oh, I can tell, it's in your voice." And – I'm just like, "Okay." (Ty S8|W44)

By the last session, only 5/18 participants (Gareth, Kevin, Panda, Southern Boy, and Elias) agreed with Q43; this was consistent since the first session for all but Kevin and Panda, who were initially ambivalent but whose month-to-month responses show a gradual shift to agreement. Only 4/18 (Chimmy, Renaud, Saint, and Wizard) were ambivalent; while this was consistent for Chimmy and Renaud, both Saint and Wizard had initially strongly agreed, revealing a shift to ambivalence. Note that by the final session, Saint and Renaud are the only participants who referred to themselves as *straight* (though Saint also used *gay*, see §3.4.2), so again we see, quite logically, why being perceived as *not straight* may not be especially desired. This patterns more closely with the majority of speakers; in addition to Marshmallow and Ty, the remaining six (ATM, Arrow, Freddy, Kel, Lizard, and Oz) all expressed that they would prefer not to be assumed *not straight* by their voice.

6.6 Conclusion

Speakers' overall satisfaction with their voices improved over time on HRT, as predicted; however, individual speakers expressed different lingering concerns, including dysphoria, feeling that their voice was still "at odds" with their gender identity after a year, and for several who desired it, issues passing or being read as male.

Satisfaction with one's voice may be contextual, or predicated on different things across speakers; Raine White did not dislike his voice on its own, but did dislike how others assumed he was female because of it. In contrast, Gareth's comfort depended entirely on how they sounded to themselves, and while ultimately happy with their voice, the process involved feelings of claustrophobia (which may be related to Gareth's reduced vocal range) and mourning for their previous voice, which they had by then spent many decades cultivating and coming to terms with. The sensation of unfamiliarity and strangeness regarding the changed voice can be a source of some distress, even as the speaker welcomes it. No speaker regretted their vocal change, and most hoped that it would continue to change, but it remains a major shift requiring a complicated adjustment and periods of, potentially, not recognizing oneself in recordings. Combined with the issues with singing described in §4.3.3, which appear less the result of decreased pitch range and more of a need to relearn muscle memory and a general issue of vocal strength and projection, it is expected that even participants showing high satisfaction with their voices will have, effectively, an adjustment period.

Time on HRT did reduce vocal dysphoria overall, but many participants reported still experiencing dysphoria by the end of the study, including some who otherwise expressed high satisfaction; voice change appears likely to not lead to a complete elimination of dysphoria. Speakers did not *consciously* change in their attempts to speak in any particularly-gendered style over the course of the year, save for those who'd adopted and then abandoned a feminine speaking style for safety, though there was a trend towards reduced efforts to intentionally deepen one's voice; this is as expected now that speakers' pitches have lowered.

Participants found their voice over the year to sound more *male*, *masculine*, and *androgynous* to them, as well as less *female* and *feminine*. Self-perception of vocal gender does not necessarily match identity; several speakers who did not call themselves *male* nonetheless identified their voices as such, but not necessarily as *masculine*. Chimmy, Kevin, Oz, and Ty were the only participants who did not identify their voices as *masculine* by the end of the study; Chimmy notably refrained from using terms like *masculine* for himself throughout the study, but nonetheless did express dissatisfaction with his voice as a result of still being misgendered vocally and over the phone. Thus, low perceived masculinity in the voice was an issue, even though Chimmy did not consider *himself* necessarily masculine.

Many speakers identified their voice as mixed-gender, i.e. as both *feminine* and *masculine* or both *female* and *male*; while this was common among non-binary speakers, it was also seen in several who were not non-binary, including Lizard and Marshmallow, and other non-binary speakers existed who viewed their voice only as some combination of *male/masculine/androgynous*. This supports what has been previously stated: that perception of one's voice, like identity, does not treat *male* and *female* as opposite ends of a scale. This has implications particularly for the analysis of acoustic data, where lower F0 is near-universally viewed as "more male" or "more masculine" than higher F0, and so on.

Speakers generally cared how their voice gendered them to others, and a mismatch in this regard corresponded with expressions of satisfaction; others' perceptions of one's voice's gender changed in roughly the same pattern as self-perception. Some non-binary speakers comfortably passed as male, others struggled with being read as male and with the impossibility of passing as non-binary, and still others eschewed the concept of passing altogether. Though perceived androgyny in the voice generally increased, this was desired by some speakers and undesired by others; ATM was frustrated by others realizing that he was not a cis woman, and insisting upon using a gender-neutral form of his name. Thus gender neutrality can itself, for a given speaker, be a form of misgendering.

In-person misgendering was reduced but continued for many speakers at the end of the study, including in participants who felt that their *voices* were perceived as male; i.e. voice was insufficient to counteract whatever physical qualities led to others misgendering them in-person. Over the phone, most speakers saw similar improvement, through some such as Freddy and Chimmy had jobs that required heavy phone usage and were subjected to substantially more of this than others. While Freddy by the final session was gendered correctly on the phone the vast majority of the time, Chimmy was not, having only begun to no longer be "ma'am'd" after beginning vocal therapy. Participants saw a reduction in desire for their voices to sound more *masculine*, *male*, and *androgynous*, though individuals expressed a common sentiment of hoping for additional vocal change. Some who found themselves embracing a more feminine identity or presentation (now that the effects of HRT have led to this not necessarily leading to misgendering) expressed the wish for their voice to sound more *feminine*, though never more *female*. There was no change over time in speakers' desire to pass as a cisgender man.

Overall, while satisfaction typically increased, participants had individual issues ranging from misgendering over the phone to continued dysphoria, failure to pass as cis when desired, or being read as vocally *androgynous* rather than *male*. 8/18 speakers indicated a desire for additional vocal change, and of them 6 also were considering, or at the time already in, vocal therapy (along with three others speakers ambivalent about desiring further change). I am glad to find that the speakers who do not feel they are entirely happy with the results of HRT are largely also those interested in vocal therapy, as this suggests a greater awareness and accessibility of vocal therapy for transmasculine speakers than has historically been available, and also that speakers who have some remaining vocal dissatisfaction are not under the impression that there is nothing that they can do to improve it.

It is my hope that his work has brought awareness to the kinds of vocal dissatisfaction still found in speakers after the first year of HRT. While these participants were, ultimately,

relatively satisfied with their voices, those developing vocal therapy techniques (or those speakers without access to therapy who wish to work to improve satisfaction their own), would do well to take many of the points here into consideration. Future work on vocal satisfaction should do the same; as seen in multiple cases above, this satisfaction questionnaire (and any such quantitative metric) inherently simplifies the problem. Effectively, it sacrifices our ability to understand the situations of individuals in favor of a broader scope.

The latter has its uses; it is no small claim that transmasculine individuals tend to be more comfortable with themselves, and less vocally dysphoric, after a year of HRT. However, any work towards *reduction* of the dissatisfaction rates cited in other studies (e.g. 12-17% from van Borsel et al. 2000, with $\geq 25\%$ still read as female over the phone and $\geq 25\%$ pursuing vocal therapy) must take into account the goals and identities of each speaker specifically. I cannot provide a comprehensive percentage of "dissatisfied" speakers here. I can say that 39% were still misgendered over the phone, that 50% desired vocal therapy, or that 28% did not vocally pass as cisgender men. I can also say that 17% still found their voices "at odds" with their gender identity, and that 28% showed persistent vocal dysphoria. 89% generally liked the sound of their voice; the latter, perhaps, is closest to how satisfaction is discussed in current literature. Each of these is a useful metric, each a simplification in its own way.

The next step is to combine our understanding of the acoustics of vocal change on HRT and of the myriad ways of understanding satisfaction in the context of individual goals and identity. While this combined analysis itself will appear in forthcoming work, we may conclude with a discussion of what we might ask, and what we might expect to find.

CHAPTER 7

CONCLUSION AND FUTURE WORK

Even though there's a lot of toxicity in the world, there's also a lot of real progress. And, um, I think it's gonna be really fascinating to see [...] if our culture around this has changed so much, just in the last ten or twenty years, what are the next ten or twenty years going to bring? (Gareth S10|W47)

7.1 Overview of results

This remote, longitudinal study made use of monthly interviews and a series of acoustic tasks, including a read passage, image description, and pitch range task, in order to measure properties of pitch and the vowel space over the course of 18 transmasculine English speakers' first year of HRT. Participants were varied in age, race, identity, sexuality, and neurodiversity; 11/18 used the term *non-binary* to refer to themselves, while 14/18 used *trans man*, illustrating an overlap of identities between which there is no clear dividing line.

Here I summarize the results of each element of data collection, before concluding with methodological concerns and a look towards future work.

7.1.1 Pitch

In pitch, participants saw the expected (Deuster et al., 2016; Nygren et al., 2009; Irwig et al., 2017) lowering of FO across all areas of measurement: habitual pitch, ceiling, and floor. All speakers' final habitual pitches were below the "gender ambiguous" 185Hz target pitch used by Ziegler et al. (2018); only 61% were below the "cisgender male normative" 131Hz, but the use of these metrics is inherently arbitrary and should not be taken as indicative of "poor" results. Participants saw several patterns of change in habitual pitch and floor; most showed the expected "cliff-and-plateau" pattern of drastic change in the first 13 weeks, while others

saw a delay of several months before pitch change commenced and still others saw gradual descent over the entire first year (see Irwig et al. 2017). Overall distance between speakers' habitual pitch and floor did not change, though a subset of five speakers did see an increase in this distance over time.

Most speakers had unusually narrow (≤ 24 ST) vocal ranges measured both pre- and post-HRT, suggesting some influence of study methodology on ability to produce true minimum and maximum pitches. Vocal range in Hz did narrow over time across participants, but when converted to a semitone scale this range narrowing was negligible for most speakers. Three did exhibit significant narrowing; two instead saw widening. Overall, we find no evidence that HRT has a consistent narrowing effect on the vocal range, but acknowledge that some speakers do see some extent of narrowing, which must be taken into consideration in any efforts to improve speaker satisfaction. Floor and habitual pitch typically descended in parallel, but not for all speakers; some instead found floor to drop sharply, with habitual pitch taking time to descend and match it. Most speakers saw significant lowering of pitch ceiling, but some maintained access to their upper register, and several developed access to falsetto that regained part of that range over time. Many singers expressed a newfound enjoyment of singing with their changed voices, but also frustration with vocal cracks, hoarseness, or most commonly a sense of inability to pitch-match; the muscle memory of a note no longer corresponds to the note produced. Thus, even without a narrowing of range, singers on testosterone HRT can expect an adjustment period.

These elements of vocal health, such as cracks, hoarseness, sore throat, difficulty projecting or speaking loudly, etc. are not considered here; they were discussed qualitatively by individual speakers, and will be addressed in future work examining speaker interviews.

7.1.2 Vowel Characteristics

In vowels, we find speakers generally exhibiting lowered formant frequencies in one or both of F1 and F2 over time, as well as raising in F3, with the latter change the strongest and most consistent; these results are comparable to those of Papp (2012) and Zimman (2012). These changes, however, are uniform across neither participants nor vowels; different speakers show different patterns of change, including not only raising in one formant and lowering in another but raising in one *vowel* in one formant, and lowering in another.

No conclusive patterns could be found in natural classes among the vowels that tended to exhibit the most change, though lowering may have been more common in close and open front vowels. Vocal tract length (VTL) was estimated for all speakers using F3, showing no change for most speakers and a decrease for four; this corresponds to the F3 raising seen, but is unexpected given the prediction (Cler et al., 2020; Hodges-Simeon et al., 2021; Buckley et al., 2022) that HRT may lengthen the vocal tract. The change in habitual pitch, floor, and ceiling are to be expected given the physiological effects of testosterone. The participant-specific effects on mean-floor distance and vowel formant frequencies are not. I suggest these latter to be primarily the result of articulatory shift motivated by a combination of sociolinguistic factors (as speakers interact differently with gendered characteristics of the voice over the course of transition) and indirectly physiologically-motivated ones (as speakers adjust to their changing pitch by changing articulation, with idiosyncratic results). The raised F3 is consistent enough across speakers to suggest some common pattern, but at this time I have no hypothesis as to its origin.

Differences were also seen between elicitation tasks, both in pitch and in vowel characteristics; these differences were not consistent across speakers, and so do not support the hypothesis that a read-passage task (given evidence against its representativeness of conversational speech (Howell and Kadi-Hanifi, 1991; Zraick et al., 2000; Nakamura et al., 2008; Schiwitz, 2011) as well as the salience of reading aloud to trans speakers in the specific context

of voice change) would show more carefully performed, and therefore closer-to-speaker-ideal vocal characteristics. However, I hesitate to discard this hypothesis entirely. Significant cross-task differences were found for 8 participants in pitch, 5 in overall vowel formants, and 12 in F2 specifically, where the read-passage produced lower F2s. Given finding by Leung et al. (2021) that listener perception of speaker gender was predicted primarily by F2 (and F0), it is still very possible that these task effects are, at least in part, linked to situational performance of a more male-typical vocal style.

7.1.3 *Satisfaction*

Overall satisfaction improved over time on HRT, dysphoria reduced, and individuals reported their voices to be more *male*, *masculine*, and *androgynous* both in self-perception and in their understanding of others' perceptions of them. Happiness with their voices increased, disconnect between voice and gender identity decreased, and in general participants were extremely welcoming of the changes experienced. However, satisfaction is a nebulous concept, and individuals did show numerous specific issues by the end of the study; many retained some vocal dysphoria, did not consistently pass as male in public (and wanted to do so), or found their voices undesirably androgynous. Misgendering over the phone was a frequent problem. Many pursued, or reported intent to pursue, vocal therapy during or following the study, in order to continue to develop their voices now that HRT has made the majority of its expected vocal changes.

Though HRT had an overwhelmingly positive effect on these participants' vocal situations, they were not without dissatisfaction, and the nature of that dissatisfaction was informed largely by their individual vocal goals and identities; one man's *frustratingly androgynous* is another man's *man*.

7.2 Methodological considerations

There are some methodological concerns to address; most notably, as this work began during the COVID-19 pandemic, it was performed entirely remotely. §3 describes the efforts made to ensure the cleanest possible recordings for acoustic analysis, but remote phonetics work remains relatively new. We cannot ignore the possibility of confounding data having been introduced via noise, compression, or differences in audio quality.

Secondly, the questionnaires used here were compiled from an array of existing metrics (primarily the TVQ and TSEQ) as well as through informal collaboration with other trans individuals to develop a list of questions and terms that they would ask themselves or use. The VENI (Voice-related Experiences of Nonbinary Individuals) questionnaire would have been appropriate to use here, but was unfortunately in development concurrently with this study's methodology and participant recruitment phase. The resulting questionnaires are not perfect; as detailed in §6.2.2, several questions had conflicting interpretations. However, I believe they did serve their purpose in prompting discussion; as I argue in §6.6, when it comes to satisfaction, though a quantitative metric serves to highlight trends, it is predominantly on the level of individual understanding that *improvements* to satisfaction can be made.

7.3 Trajectory of future work

There is substantially more to be done. First, now that acoustic change in this cohort has been examined, and various metrics of satisfaction both quantified and given some qualitative context, we may ask what, if any, are acoustic correlates of satisfaction for these speakers.

Nygren et al. (2009) found speaker satisfaction to correlate with lower F0; Deuster et al. (2016) found the *extent* of F0 change to be more predictive than final F0 itself. I hypothesize that change in vocal range will also be relevant here, for those who saw it reduce. I also suggest particular attention be given to vowel characteristics. Given known gendered features

of the vowel space (Canessa-Pollard et al., 2012; Gelfer and Bennett, 2013; Vorperian et al., 2019), the emphasis placed on "resonance" in trans spaces and vocal therapy contexts as a method of manipulating perceived vocal gender, possible correlations between resonance and the vowel space (Cai, 2019), and Zimman (2012)'s suggestion that the explanation for the characteristic "transmasculine accent" lies in the vowel space, there is no doubt that the vowel spaces of transmasculine speakers are worth a closer look. F3 in particular bears investigation; this raising is not unique to this study, but directly contradicts other findings that HRT increases vocal tract length, which we would expect to have a uniform effect on all formants. What, we may ask, causes this dramatic increase in F3, and do we see different results for estimated VTL when taking F4 into account?

We may also look at other gender-associated features of the voice as described in §2.2.5, such as properties of /s/ and /ʃ/, use of creak and other non-modal voices, and intonation. Intonation in particular is highly salient to speakers; there is evidence of some difference in intonation between genders, with recent work even specifically examining non-binary speech (Hope and Bradley, 2019). Notably, there is also evidence that even where intonational patterns do *not* differ between genders Henton (1995), the cultural belief that they do can itself be useful to trans speakers. Trans women who incorporate greater pitch dynamics appear more likely to be read as women (Wolfe et al., 1990), and so it is possible that the old transmasculine passing tip of "speak in a monotone" has some truth to it, even if it is not wholly representative of cis male speech.¹

Features of fricatives are of particular note due to their association with other elements of LGBTQ language, and gay male speech in particular (Munson et al., 2006; Mack and Munson, 2012). The degree to which participants in the current study expressed a desire to be read as non-straight by others is notable, as we may expect those who identify as gay

1. Should I ever delve into the realm of prosody, I would be eager to look into the use of high rising terminal in trans speakers, not only because it was the single most-mentioned gender-associated feature other than pitch by participants in the current study, but for the delight of getting to study "HRT on HRT."

men to incorporate some elements of these speech patterns. However, there is also evidence that the transmasculine speech *already* incorporates some formant characteristics also found in gay male speech, and that various acoustic properties and spectral moments of these fricatives may in fact change over time on HRT (Papp, 2012; Zimman, 2012).

Next, we may ask about the intersections between trans experiences and multilingualism; the gendered characteristics discussed in this work are extremely English-specific, and as pointed out in §2.1.1, even something as clearly biologically defined as pitch, on which HRT is known to have an effect, remains at the whims of sociolinguistic factors as well. Habitual pitches are not consistent across languages; how and whether trans speakers' voices adapt to this is worth consideration, particularly in the case of multilingual trans speakers who navigate different, using different languages, different social contexts – which may themselves come with different expectations of gender, presentation, and identity. This is a question we could take some initial steps towards through this study; many participants were bilingual, and in addition to the vocal tasks performed as part of the interviews, several of them provided additional recordings in other languages spoken.

There is also the possibility of additional sessions with these same participants, positioned one or two years after the end of the study. It is very likely that any biological change caused by HRT in the vocal tract will grow more prominent over time², and there is relatively little work done on acoustic change beyond the first year of HRT, particularly work incorporating qualitative discussion. Many participants have already indicated they would be amenable to an additional session in the coming year; should we see different results in estimated VTL or formants after this time, that could suggest there is a physiological effect of HRT on the vowel space that simply does not apply uniformly within the first year.

Perceptual work would also be worthwhile, particularly following up on the issues raised regarding "external perception" and how it is determined. We may ask to what extent

2. I suggest in §5.3 that this is a potential source of the VTL change discrepancy with Hodges-Simeon et al. (2021).

speakers' impressions of how others gender their voices are in fact reflective of how others do so, and perform a three-way comparison with how *speakers* do so. This would allow us not only to gauge how accurate speakers are at inferring this, but to determine whether one or more of these corners of perception have a great impact on speaker vocal satisfaction.

Finally, there is another topic of satisfaction worth discussion, involving the two speakers who, despite their participation throughout this study, were not ultimately included in the analysis. Lemon and the Dark Lord (their chosen aliases) both began HRT but were unable to remain on it for the duration of the year; both only began again in the study's final weeks. As a result, they showed minimal vocal change, and their acoustic data was omitted; however, their *satisfaction* data, not only the questionnaires but the conversations and qualitative interview responses, bear acknowledgement. It is not uncommon for trans people to be unable to access medical transition, either due to a lack of resources, funds, medical contraindications, or living where such access is prohibited. My interviews with these participants were just as insightful as the rest, and furthermore there were substantial components to all participants' experiences, relevant to their personal satisfaction with their voices, that for the sake of space, time, and coherency I could not include here.

We may follow up the acoustic data with further analysis, and comparison with the quantified satisfaction results; we may follow the qualitative satisfaction results with a deeper look into these interviews, via a content analysis in the manner of Azul (2015a). Ultimately, I intend to combine the two, allowing for insight into how the vast array of components which have the potential to affect satisfaction³ have impacted these speakers – and by extension, many others. This may also incorporate a discussion of vocal health, in terms of reported creaking, hoarseness, strain, or the like, which may have great impact on satisfaction but which was omitted here as it was not strictly quantified.

3. Habitual pitch, floor, ceiling, range, vowel formants, VTL, resonance, features of fricatives, gender identity, sexuality, presentation, environment, intonation, the "transmasculine accent", intentional use of gendered speech styles, voice training, self-perception of the voice, external perception of the voice, physical vocal comfort, and so on, just to list a few.

7.4 Final comments

This dissertation is preliminary, and not without flaws. It asks many questions, and offers few answers; in the end, much of my conclusion must amount to an insistence of the acknowledgement of the sheer breadth of individual variation in transmasculine speakers, and a warning against assuming that what leads to greater vocal satisfaction for one will necessarily hold true for another. Regardless, this remains a contribution to our understanding of trans vocal change; building on prior longitudinal studies, it confirms and clarifies some patterns long suggested, including rising F3 and the lack (for most speakers) of range narrowing. It critiques the read passage task's usefulness in evaluating trans voices, and acts as a pilot study for image description as an alternate method; though the degree of speaker-specific variation between these tasks leaves results uncertain, I argue that we must still be cautious in the assumption that read speech is an accurate metric of trans voices.

Ultimately, to make headway in improving satisfaction for trans speakers, we must not only develop a firm understanding of how acoustic properties of the voice relate to perception of gender, how they are affected by HRT, and how personal identity and behavioral articulatory change play a role. We must then work with vocal therapists, and with trans speakers themselves, to develop ways of implementing this understanding into practical advice. After approximately 12 months of HRT, 89% of this cohort felt that they "sounded like themselves". It is my hope that through the rapidly-growing body of research to which this work contributes, the same may one day be said for us all.

APPENDIX A

RAINBOW PASSAGE

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon. There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.

Throughout the centuries, people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation. To the Hebrews, it was a token that there would be no more universal floods. The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain. The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their home in the sky. Others have tried to explain the phenomenon physically. Aristotle thought that the rainbow was caused by reflection of the sun's rays by the rain. Since then, physicists have found that it is not reflection, but refraction by the raindrops which causes the rainbows.

Many complicated ideas about the rainbow have been formed. The difference in the rainbow depends considerably upon the size of the drops, and the width of the colored band increases as the size of the drops increases. The actual primary rainbow observed is said to be the effect of super-imposition of a number of bows. If the red of the second bow falls upon the green of the first, the result is to give a bow with an abnormally wide yellow band, since red and green light when mixed form yellow. This is a very common type of bow, one showing mainly red and yellow, with little or no green or blue.

REFERENCES

- Abelson, M. (2016). ‘You aren’t from around here’: Race, masculinity, and rural transgender men. *Gender, Place & Culture*, 23:1–12.
- Adank, P., Smits, R., and van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *J. Acoust. Soc. Am.*, 116:3099–3107.
- Adler, R. K., Hirsch, S., and Pickering, J. (2019). *Voice and communication therapy for the transgender/gender diverse client: A comprehensive clinical guide*. Plural Publishing, Inc., San Diego, CA, third edition.
- Alves, M., Krüger, E., Pillay, B., van Lierde, K., and van der Linde, J. (2019). The effect of hydration on voice quality in adults: A systematic review. *J. Voice*, 125:125.e13–125.e28.
- American Psychiatric Association (2022). *Diagnostic and statistical manual of mental disorders (5th ed., text rev.)*.
- Amir, O., Amir, N., and Michaeli, O. (2005). Evaluating the influence of warmup on singing voice quality using acoustic measures. *J. Voice*, 19(8):256–260.
- Anderson, A. D., Irwin, J. A., Brown, A. M., and Grala, C. L. (2020). “Your picture looks the same as my picture”: An examination of passing in transgender communities. *Gender Issues*, 37(1):44–60.
- Auer, M., Fuss, J., Höhne, N., Stalla, G., and Sievers, C. (2014). Transgender transitioning and change of self-reported sexual orientation. *PLoS One*, 9(10):e110016.
- Azul, D. (2015a). Gender-related aspects of transmasculine people’s vocal situations: insights from a qualitative content analysis of interview transcripts. *Int. J. Lang. Commun. Disord.*, 56(1):672–684.
- Azul, D. (2015b). Transmasculine people’s vocal situations: a critical review of gender-related discourses and empirical data. *Int. J. Lang. Commun. Disord.*, 50:31–47.
- Azul, D., Hancock, A., and Nygren, U. (2021). Forces affecting voice function in gender diverse people assigned female at birth. *J. Voice*, 35.
- Azul, D., Nygren, U., Södersten, M., and Neuschaefer-Rube, C. (2017). Transmasculine people’s voice function: A review of the currently available evidence. *J. Voice*, 31(2)(2):261.e9–261.e23.
- Bachorowski, J.-A. and Owren, M. (1999). Acoustic correlates of talker sex and individual talker identity are present in short vowel segment produced in running speech. *J. Acoust. Soc. Am.*, 106:1054–63.
- Baken, R. and Orlikoff, F. (2000). *Clinical Measurement of Speech and Voice (2nd ed.)*. Singular Publishing, San Diego CA.

- Becker, K., Khan, S. D., and Zimman, L. (2022). Beyond binary gender: Creaky voice, gender, and the variationist enterprise. *Language Variation and Change*, 34(2):215–238.
- Beischel, W., Gauvin, S., and van Anders, S. (2021). "A little shiny gender breakthrough": Community understandings of gender euphoria. *Int. J. Trans. Health*, 23(3).
- Benjamin, H. (1967). The transsexual phenomenon*. *Transactions of the New York Academy of Sciences*, 29(4 Series II):428–430.
- Block, C. (2017). Making a case for transmasculine voice and communication training. *Perspectives of the ASHA Special Interest Groups*, 2 (SIG3):33–41.
- Boersma, P. and Weenink, D. (2023). Praat: Doing Phonetics by Computer, Version 6.3.09.
- Booz, J. A. (2016). Perceived gender in clear and conversational speech. *J. Acoust. Soc. Am.*, 139:2107–2107.
- Bucholtz, M. and Hall, K. (2004). Theorizing identity in language and sexuality research. *Language in Society*, 33(4):469–515.
- Bucholtz, M. and Hall, K. (2005). Identity and interaction: a sociocultural linguistic approach. *Discourse Studies*, 7(4-5):585–614.
- Buckley, D., Dahl, K., and Stepp, C. (2022). Transmasculine voice modification: A case study. *J. Voice*, 34(6).
- Bultynck, C., Pas, C., Defreyne, J., Cosyns, M., den Heijer, M., and T'Sjoen, G. (2017). Self-perception of voice in transgender persons during cross-sex hormone therapy. *The Laryngoscope*, 127(12).
- Bultynck, C., Pas, C., Defreyne, J., Cosyns, M., and T'Sjoen, G. (2020). Organizing the voice questionnaire for transgender persons. *Int. J. Trans. Health*, 21(1):89–97.
- Bustos, V. P., Bustos, S. S., Mascaro, A., Del Corral, G., Forte, A. J., Ciudad, P., Kim, E. A., Langstein, H. N., and Manrique, O. J. (2021). Regret after gender-affirmation surgery: A systematic review and meta-analysis of prevalence. *Plastic and reconstructive surgery. Global open*, 9(3).
- Böhme, G. (1997). *Sprach-, Sprech-, Stimm- und Schluckstörungen: Ein Lehrbuch*. Gustav Fischer, Stuttgart.
- Cai, H. (2019). Acoustic analysis of resonance characteristics of head voice and chest voice. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE.
- Cai, X., Hughto, J., Reisner, S., Pachankis, J., and Levy, B. (2018). Benefit of gender-affirming medical treatment for transgender elders: Later-life alignment of mind and body. *LGBT Health*, 6.

- Calder, J. and King, S. (2020). Intersections between race, place, and gender in the production of /s/. *University of Pennsylvania Working Papers in Linguistics*, 26(2).
- Califia, P. (1997). *Sex Changes: The Politics of Transgenderism*. Cleis Press, San Francisco, CA, first edition.
- Canessa-Pollard, V., Cowles, H., and Reby, D. (2012). Spontaneous voice gender imitation abilities in adult speakers. *PLoS One*, 7.
- Cler, G. J., McKenna, V. S., Dahl, K. L., and Stepp, C. E. (2020). Longitudinal case study of transgender voice changes under testosterone hormone therapy. *J. Voice*, 34(5):748–762.
- Coleman, E., Radix, A. E., Bouman, W. P., Brown, G. R., de Vries, A. L. C., Deutsch, M. B., Ettner, R., Fraser, L., Goodman, M., Green, J., Hancock, A. B., Johnson, T. W., Karasic, D. H., Knudson, G. A., Leibowitz, S. F., Meyer-Bahlburg, H. F. L., Monstrey, S. J., Motmans, J., Nahata, L., Nieder, T. O., Reisner, S. L., Richards, C., Schechter, L. S., Tangpricha, V., Tishelman, A., Trotsenburg, M. A. A. V., Winter, S., Ducheny, K., Adams, N. J., Adrián, T., Allen, L. R., Azul, D., Bagga, H., Başar, K., Bathory, D. S., Belinky, J. J., Berg, D., Berli, J. U., Bluebond-Langner, R. O., Bouman, M.-B., Bowers, M. L., Brassard, P. J., Byrne, J., Capitán, L., Cargill, C. J., Carswell, J. M., Chang, S. C., Chelvakumar, G., Corneil, T., Dalke, K. B., Cuypere, G. D., de Vries, E., Heijer, M. D., Devor, A. H., Dhejne, C., D’Marco, A., Edmiston, E. K., Edwards-Leeper, L., Ehrbar, R., Ehrensaft, D., Eisfeld, J., Elaut, E., Erickson-Schroth, L., Feldman, J. L., Fisher, A. D., Garcia, M. M., Gijs, L., Green, S. E., Hall, B. P., Hardy, T. L. D., Irwig, M. S., Jacobs, L. A., Janssen, A. C., Johnson, K., Klink, D. T., Kreukels, B. P. C., Kuper, L. E., Kvach, E. J., Malouf, M. A., Massey, R., Mazur, T., McLachlan, C., Morrison, S. D., Mosser, S. W., Neira, P. M., Nygren, U., Oates, J. M., Obedin-Maliver, J., Pagkalos, G., Patton, J., Phanuphak, N., Rachlin, K., Reed, T., Rider, G. N., Ristori, J., Robbins-Cherry, S., Roberts, S. A., Rodriguez-Wallberg, K. A., Rosenthal, S. M., Sabir, K., Safer, J. D., Scheim, A. I., Seal, L. J., Sehoole, T. J., Spencer, K., Amand, C. S., Steensma, T. D., Strang, J. F., Taylor, G., Tilleman, K., T’Sjoen, G. G., Vala, L. N., Mello, N. M. V., Veale, J. F., Vencill, J., Vincent, B., Wesp, L. M., West, M. A., and Arcelus, J. (2022). Standards of care for the health of transgender and gender diverse people, version 8. *Int. J. Trans. Health*, 23(sup1):S1–S259.
- Coleman, R. and Markham, I. (1991). Normal variations in habitual pitch. *J. Voice*, 5(2):173–177.
- Coleman, R. O. (1983). Acoustic correlates of speaker sex identification: Implications for the transsexual voice. *J. Sex Research*, 19(3):293–295.
- Colton, R., Casper, J., and Leonard, R. (2011). *Understanding voice problem: A physiological perspective for diagnosis and treatment*. Fourth edition.
- Constansis, A. N. (2008). The changing female-to-male (FTM) voice. *Radical Musicology*, 3.

- Cooksey, J. (2000). Voice transition in male adolescents. In Thurman, L. and Welch, G., editors, *Bodymind and Voice: Foundation of Voice Education*, pages 717–738. Voicecare Network, Iowa City, IA.
- Cosyns, M., Borsel, J. V., Wierckx, K., Dedecker, D., de Peer, F. V., Daelman, T., Laenen, S., and T'Sjoen, G. (2014). Voice in female-to-male transsexual persons after long-term androgen therapy. *Laryngoscope*, 124(6):1409–1414.
- Dacakis, G., Davies, S., Oates, J., Douglas, J., and Johnston, J. (2013). Development and preliminary evaluation of the transsexual voice questionnaire for male-to-female transsexuals. *J. Voice*, 27.
- Dahl, K., François, F., Buckley, D., and Stepp, C. (2022). Voice and speech changes in transmasculine individuals following circumlaryngeal massage and laryngeal reposturing. *Am. J. Speech*, 31(3):1368–1382.
- Damrose, E. (2009). Quantifying the impact of androgen therapy on the female larynx. *Auris Nasus Larynx*, 36:110–112.
- Davies, S. (2006). Transgender self-evaluation questionnaire. In Adler, R. K., Hirsch, S., and Mordaunt, M., editors, *Voice and Communication Therapy for The Transgender/Transsexual Client: A Comprehensive Clinical Guide*, pages 485–487. Plural Publishing, San Diego, CA.
- Davies, S. and Goldberg, J. (2006). Clinical aspects of transgender speech feminization and masculinization. *Int. J. Transgenderism*, 9:167–196.
- Davies, S., Papp, V. G., and Antoni, C. (2015). Voice and communication change for gender nonconforming individuals: Giving voice to the person inside. *Int. J. Transgenderism*, 16(3):117–159.
- De Decker, P. and Nycz, J. (2011). For the record: Which digital media can be used for sociophonetic analysis? *University of Pennsylvania Working Papers in Linguistics*, 17(2):51–59.
- de Vries, K. and Sojka, C. (2020). Transitioning gender, transitioning race: Transgender people and multiracial positionality. *Int. J. Trans. Health*, 23:1–11.
- Demenko, G., Möbius, B., and Andreeva, B. (2014). Analysis of pitch profiles in Germanic and Slavic languages. In *Forum Acusticum 2014, Krakow, Poland*.
- Descloux, P., Isoard-Nectoux, S., Matoso, B., Matthieu-Bourdeau, L., Schneider, F., and Schweizer, V. (2012). Transsexuality: speech therapy supporting the "voice" of transformation. *Revue de laryngologie-otologie-rhinologie*, 133(1):41–44.

- Deuster, D., Vincenzo, K., Szukaj, M., Zehnhoff-Dinnesen, A., and Dobel, C. (2016). Change of speech fundamental frequency explains the satisfaction with voice in response to testosterone therapy in female-to-male gender dysphoric individuals. *European Archives of Oto-Rhino-Laryngology*, 273(8):2127–2131.
- Diamond, L. M. and Butterworth, M. (2008). Questioning gender and sexual identity: Dynamic links over time. *Sex Roles*, 59:365—376.
- Doyle, J. E. S. (February 17, 2022). The unique challenges and rewards of transitioning later in life. <https://xtramagazine.com/power/identity/transitioning-late-in-life-218262>.
- Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, 1(3):245–267.
- Eichhorn, J. T., Kent, R. D., Austin, D., and Vorperian, H. K. (2018). Effects of aging on vocal fundamental frequency and vowel formants in men and women. *J. Voice*, 32(5):644.e1–644.e9.
- Esling, J. (2005). There are no back vowels: The laryngeal articulator model. *The Canadian Journal of Linguistics / La revue canadienne de linguistique*, 50:13–44.
- Esposito, C. M. and Khan, S. D. (2020). The cross-linguistic patterns of phonation types. *Language and Linguistics Compass*, 14(12):e12392.
- Evans, S., Neave, N., Wakelin, D., and Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology Behavior*, 93(4-5)(4-5):783–788.
- ExistingCactus (2022). "Trans voice" exists because there are multiple components to how a voice sounds, and testosterone only changes one of them. [Comment on the online forum post *How do I avoid the "trans voice" once I start T? (FtM, <2 weeks until I start)*.] Reddit. https://www.reddit.com/r/asktransgender/comments/sd60ei/how_do_i_avoid_the_trans_voice_once_i_start_t_ftm/.
- Fairbanks, G. (1960). *Voice and articulation drillbook, 2nd edition*. Harper & Row, New York.
- Farrington, C., King, S., and Kohn, M. (2021). Sources of variation in the speech of African Americans: Perspectives from sociophonetics. *Wiley interdisciplinary reviews. Cognitive science*, 12(3):e1550.
- Fitch, W. and Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *J. Acous. Soc. Am.*, 106:1511–22.
- Flynn, N. (2011). Comparing vowel formant normalisation procedures. *York Working Papers in Linguistics (Series 2)*, 11:1–28.

- French, P., Saxena, M., Harrison, P., and Künzel, H. J. (1998). Normative f0 data for panjabi and urdu: preliminary report on male speakers over telephone lines. In *International Association for Forensic Phonetics and Acoustics Annual Meeting, Voorburg, Netherlands*.
- Gallena, S. J. K., Stickels, B., and Stickles, E. (2018). Gender perception after raising vowel fundamental and formant frequencies: Considerations for oral resonance research. *J. Voice*, 32(5):592–601.
- Ge, C., Xiong, Y., and Mok, P. (2021). How reliable are phonetic data collected remotely? Comparison of recording devices and environments on acoustic measurements. In *Proc. 22nd Annual Conference of the International Speech Communication Association (INTER-SPEECH)*, pages 1683–1687.
- Gelfer, M. P. and Bennett, Q. E. (2013). Speaking fundamental frequency and vowel formant frequencies: effects on perception of gender. *J. Voice*, 27(5):556–566.
- Gelfer, M. P. and Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *J. Voice*, 19(1):25–45.
- Gelfer, M. P. and Mordaunt, M. (2018). Pitch and intonation. In Adler, R. K., Hirsch, S., and Mordaunt, M., editors, *Voice and Communication Therapy for The Transgender/Transsexual Client: A Comprehensive Clinical Guide (3rd ed.)*, pages 197–224. Plural Publishing, San Diego, CA.
- Gelfer, M. P. and Schofield, K. J. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *J. Voice*, 14(1):22–33.
- Goff, P., Thomas, M., and Jackson, M. (2008). “Ain’t I a woman?”: Towards an intersectional approach to person perception and group-based harms. *Sex Roles*, 59:392–403.
- Goodglass, H. and Kaplan, E. (1983). *Boston diagnostic aphasia examination booklet*. Lea & Febiger, Philadelphia, PA.
- Gorton, R., Women, L.-M., and Buth, J. (2005). *Medical Therapy and Health Maintenance for Transgender Men: A Guide For Health Care Providers*.
- Green, J. (2020). *Navigating trans*+ and complex gender identities*. Bloomsbury Academic, London • New York.
- Grillo, E. U., Brosious, J. N., Sorrell, S. L., and Anand, S. (2016). Influence of smartphones and software on acoustic voice measures. *Int. J. of Telerehabilitation*, 54(2):292–305.
- Hancock, A., Colton, L., and Douglas, F. (2014). Intonation and gender perception: applications for transgender speakers. *J. Voice*, 28(2):203–209.

- Hancock, A. B., Childs, K. D., and Irwig, M. S. (2017). Trans male voice in the first year of testosterone therapy: Make no assumptions. *J. Speech Lang. Hear Res.*, 60(9):2472–2482.
- Hansbury, G. (2005). The middle men: An introduction to the transmasculine identities. *Studies in Gender and Sexuality*, 6(3):241–264.
- Henton, C. (1995). Cross-language variation in the vowels of female and male speakers. In *Proc. XIIIth International Congress of Phonetic Sciences*, volume 4, pages 420–423.
- Henton, C. and Bladon, A. (1988). Creak as a sociophonetic marker. In Hyman, L. M. and Li, C. N., editors, *Language, speech and mind: Studies in honour of Victoria A. Fromkin.*, London. Routledge.
- Hodges-Simeon, C. R., Grail, G. P. O., Albert, G., Groll, M. D., Stepp, C. E., Carré, J. M., and Arnocky, S. A. (2021). Testosterone therapy masculinizes speech and gender presentation in transgender men. *Scientific reports*, 11(1).
- Hope, M. and Bradley, E. D. (2019). Vocal pitch and intonation characteristics of those who are gender non-binary. In *Proc. 19th Int. Congress of Phonetic Sciences*, pages 2685–2689.
- Howell, P. and Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, 10(2):163–169.
- Irwig, M. S., Childs, K., and Hancock, A. B. (2017). Effects of testosterone on the transgender male voice. *Scientific Reports*, 5(1):107—112.
- Itize, I. R. (1994). *Principles of voice production*. PrenticeHall, Englewood Cliffs, NH.
- James, S., Herman, J., Rankin, S., Keisling, M., Mottet, L., and Anafi, M. A. (2016). The report of the 2015 us transgender survey. <http://www.ustranssurvey.org/reports>.
- Jannetts, S., Schaeffler, F., Beck, J., and Cowen, S. (2019). Assessing voice health using smartphones: Bias and random error of acoustic voice parameters captured by different smartphone types. *Int. J. Lang. and Communication Disorder*, 54(2):292—305.
- Jessen, M. (2009). *Forensic phonetics and the influence of speaking style on global measures of fundamental frequency*, pages 115–139.
- Johnson, A. H. (2019). Rejecting, reframing, and reintroducing: Trans people’s strategic engagement with the medicalisation of gender dysphoria. *Sociology of Health Illness*, 41(3).
- Kachel, S., Simpson, A. P., and Steffens, M. C. (2018). “Do I sound straight?”: Acoustic correlates of actual and perceived sexual orientation and masculinity/femininity in men’s speech. *J. Speech Lang. Hear. Res.*, 61(7):1560–1578.
- Kallitsounaki, A. and Williams, D. (2022). Autism spectrum disorder and gender dysphoria/incongruence. a systematic literature review and meta-analysis. *Journal of Autism and Developmental Disorders*, 53:1–15.

- Kapp, S. (2020). *Autistic Community and the Neurodiversity Movement Stories from the Frontline: Stories from the Frontline*.
- Kinsey, A. C., Pomery, W. B., and Martin, C. E. (2016). Kinsey scale. PsycTESTS Dataset.
- Konnolly, L. (2021). Nuance and normativity in trans linguistic research. *Journal of Language and Sexuality*, 10(1):71–82.
- Kulick, D. (1999). Transgender and language: A review of the literature and suggestions for the future. *GLQ A Journal of Lesbian and Gay Studies*, 5(4):605–622.
- Labov, W. (1981). *Field Methods of the Project on Linguistic Change and Variation*. Sociolinguistic working paper. Southwest Educational Development Laboratory.
- Labov, W., Ash, S., and Charles, B. (2005). *The Atlas of North American English: Phonetics, Phonology and Sound Change*. De Gruyter Mouton, Berlin • New York.
- Lammert, A. C. and Narayanan, S. S. (2015). On short-time estimation of vocal tract length from formant frequencies. *PLoS One*, 10(7)(9).
- Lawson, R. (2020). Language and masculinities: History, development, and future. *Annual Review of Linguistics*, 6(1):409–434.
- Leinung, M. and Joseph, J. (2020). Changing demographics in transgender individuals seeking hormonal therapy: Are trans women more common than trans men? *Transgender Health*, 5.
- Leung, Y., Oates, J., Chan, S. P., and Papp, V. (2021). Associations between speaking fundamental frequency, vowel formant frequencies, and listener perceptions of speaker gender and vocal femininity–masculinity. *Journal of Speech, Language, and Hearing Research*, 64.
- Lipski, J. M. (2008). *Varieties of Spanish in the United States*. Georgetown University Press.
- Lonc, C. (1974). Genderfuck and its delights. *Gay Sunshine*, 21:4–16.
- Mack, S. and Munson, B. (2012). The influence of /s/ quality on ratings of men’s sexual orientation: Explicit and implicit measures of the ‘gay lisp’ stereotype. *J. Phonetics*, 40(1):198–212.
- McNamara, C. (2007). Re-inhabiting an uninhabitable body: Interventions in voice production with transsexual men. *Research in Drama Education*, pages 195–206.
- Meier, S. C., Pardo, S. T., Labuski, C., and Babcock, J. (2013). Measures of clinical health among female-to-male transgender persons as a function of sexual orientation. *Arch. Sex. Behav.*, 42:463–474.
- Meyerowitz, J. (2002). *How Sex Changed: A History of Transsexuality in the United States*. Harvard University Press, Cambridge, MA.

- mrsweezydc (2022). "Dude, unlearn that term. There's no such thing as a "tranny voice." It's just an effeminate male voice that you [Comment on the online forum post *How do I avoid the "trans voice" once I start T? (FtM, <2 weeks until I start).*] Reddit. https://www.reddit.com/r/asktransgender/comments/sd60ei/how_do_i_avoid_the_trans_voice_once_i_start_t_ftm/.
- Munson, B., McDonald, E., DeBoe, N. L., and White, A. R. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *J. Phonetics*, 34(2):202–240.
- Nakamura, M., Koji, I., and Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language*, 22:171–184.
- Nitz, C. and Seek, D. (2021). Voice Pitch Analyzer. <https://voicepitchanalyzer.app/>.
- Nolan, F., McDougall, K., de Jong, G., and Hudson, T. (2006). The DyViS database: Style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *Int. J. Speech, Lang. and the Law*, 16(1):31–57.
- Nygren, U., Nordenskjöld, A., Arver, S., and Södersten, M. (2009). Effects on voice fundamental frequency and satisfaction with voice in trans men during testosterone treatment—a longitudinal study. *J. Voice*, 30(6):766.e23–766.e34.
- Oates, J. and Dacakis, G. (1997). *Venereology*, 10(3):178–87.
- Owen, K. and Hancock, A. B. (2011). The role of self- and listener perceptions of femininity in voice therapy. *Int. J. Transgenderism*, 12(4):272–284.
- Papp, V. (2012). The female-to-male transsexual voice: Physiology vs. performance in production. Dissertation, Rice University.
- Pearce, J. (2019). Identity, socialization and environment in transgender speakers: Sociophonetic variation in creak and /s/. In *Proc. 19th Int. Congress of Phonetic Sciences*, pages 29–33.
- Perry, T., Ohde, R., and Ashmead, D. (2001). The acoustic bases for gender identification from children's voice. *The Journal of the Acoustical Society of America*, 109:2988–98.
- Pettit, J. (2004). Transsexualism and sex reassignment: speech differences. In Kent, R. D., editor, *The MIT encyclopedia of communication disorders*. MIT Press.
- Podesva, R. (2014). Gender and the social meaning of non-modal phonation types. *Annual Meeting of the Berkeley Linguistics Society*, 37.
- Podesva, R. J. and Van Hofwegen, J. (2015). /s/exuality in Smalltown California: Gender Normativity and the Acoustic Realization of /s/. In *Language, Sexuality, and Power: Studies in Intersectional Sociolinguistics*. Oxford University Press.

- Quinn, S., Oates, J., and Dacakis, G. (2021). Perceived Gender and Client Satisfaction in Transgender Voice Work: Comparing Self and Listener Rating Scales across a Training Program. *Folia Phoniatrica et Logopaedica*, 74(5):364–379.
- Saito, Y., Nakamura, K., Itani, S., and Tsukahara, K. (2018). Type 3 thyroplasty for a patient with female-to-male gender identity disorder. *Case Reports in Otolaryngology*, pages Article ID 4280381, 4 pages.
- Schilt, K. and Lagos, D. (2017). The development of transgender studies in sociology. *Annual Review of Sociology*, 43:425–443.
- Schwitz, A. (2011). Effect of task type on speaking fundamental frequency in women. Master’s thesis, Auburn University.
- Shefcik, G. and Tsai, P.-T. (2023). Voice-related experiences of nonbinary individuals (VENI) development and content validity. *J. Voice*, 37(2):294.e5–294.e13.
- Sirin, S., Polat, A., and Alioglu, F. (2021). Psychometric evaluation of adapted transsexual voice questionnaire for turkish trans male individuals. *J. Voice*, 35(5):805.e27–805.e32.
- Spencer, L. E. (1988). Speech characteristics of male-to-female transsexuals: A perceptual and acoustic study. *Folia phoniatrica et logopaedica*, 40(1):31–42.
- sspencerj (2020). "i have seen that people who take higher doses, end up having a lot of voice cracks and possibly having [Comment on the online forum post *How to avoid the “trans voice”?*] Reddit. https://www.reddit.com/r/transmanlifehacks/comments/io5t7n/how_to_avoid_the_trans_voice/.
- Strang, J. F., Powers, M. D., Knauss, M., Sibarium, E., Leibowitz, S. F., Kenworthy, L., Sadikova, E., Wyss, S., Willing, L., Caplan, R., Pervez, N., Nowak, J., Gohari, D., Gomez-Lobo, V., Call, D., and Anthony, L. G. (2018). “They thought it was an obsession”: Trajectories and perspectives of autistic transgender and gender-diverse adolescents. *J. Autism Dev. Disord.*, 48(12):4039–4055.
- Sundberg, J. and Högset, C. (2001). Voice source differences between falsetto and modal registers in counter tenors, tenors and baritones. *Logopedics, Phoniatics, Vocology*, 26:26–36.
- Swaab, D. F. (2007). Sexual differentiation of the brain and behavior. *Best practice & research clinical endocrinology & metabolism*, 21(3):431–444.
- Söderpalm, E., Larsson, A. K., and Almquist, S.-A. (2004). Evaluation of a consecutive group of transsexual individuals referred for vocal intervention in the west of Sweden. *Logopedics, Phoniatics, Vocology*, 29:18–30.
- Turban, J. L., King, D., Kobe, J., Reisner, S. L., and Keuroghlian, A. S. (2022). Access to gender-affirming hormones during adolescence and mental health outcomes among transgender adults. *PLoS One*, 17(1):1–15.

- Turner, R. E., Walters, T. C., Monaghan, J. J. M., and Patterson, R. D. (2009). A statistical, formant-pattern model for segregating vowel type and vocal-tract length in developmental formant data. *J. Acoust. Soc. Am.*, 125(4):2374–2386.
- T’Sjoen, G., De Cuyper, G., Monstrey, S., Hoebeke, P., Freedman, F. K., Appari, M., Holterhus, P.-M., Van Borsel, J., and Cools, M. (2011). Male gender identity in complete androgen insensitivity syndrome. *Archives of Sexual Behavior*, 40:635–638.
- Valentine, D. (2007). *Imagining Transgender: An Ethnography of a Category*. Duke University Press, Durham, NC.
- van Bezooijen, R. (1996). The effect of pitch on the attribution of gender-related personality traits. In Warner, N., Ahlers, J., Bilmes, L., Oliver, M., Wertheim, S. W., and Chen, M., editors, *Gender and Belief Systems: Proceedings of the Fourth Berkeley Women and Language Conference.*, pages 755–766, Berkeley. Berkeley Women and Language Group.
- van Borsel, J., de Cuyper, G., Rubens, R., and Destaecke, B. (2000). Voice problems in female-to-male transsexuals. *Int. J. Language and Communication Disorders*, 35(3):427–442.
- Vorperian, H. K., Kent, R. D., Lee, Y., and Bolt, D. M. (2019). Corner vowels in males and females ages 4 to 20 years: Fundamental and f1-f4 formant frequencies. *J. Acoust. Soc. Am.*, 146(5):3255–3274.
- Webb, H., Free, N., Oates, J., and Paddle, P. (2022). The use of vocal fold injection augmentation in a transmasculine patient unsatisfied with voice following testosterone therapy and voice training. *J. Voice*, 36(4):588.e1–588.e6.
- Weirich, M. and Simpson, A. P. (2018). Gender identity is indexed and perceived in speech. *PLoS One*, 13(12):e0209226.
- Whiteside, S., Hodgson, C., and Tapster, C. (2002). Vocal characteristics in pre-adolescent and adolescent children: a longitudinal study. *Logopedics, Phoniatrics, Vocology*, 27(1):12–20.
- Whiteside, S. and Marshall, J. (2001). Developmental trends in voice onset time: Some evidence for sex differences. *Phonetica*, 58:196–210.
- Wolfe, V. I., Ratusnik, D. L., Smith, F. H., and Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *J. Speech and Hearing Disorders*, 55(1):43–50.
- Yuasa, I. P. (2008). *Culture and Gender of Voice Pitch: A Sociophonetic Comparison of the Japanese and Americans*. Equinox, London.
- Yuasa, I. P. (2010). Creaky Voice: A New Feminine Voice Quality for Young Urban-Oriented Upwardly Mobile American Women? *American Speech*, 85(3):315–337.

- Zhang, C., Jepson, K., Lohfink, G., and Arvaniti, A. (2016). Speech data collection at a distance: Comparing the reliability of acoustic cues across homemade recordings. *J. Acoust. Soc. Am.*, 148(4):2717.
- Ziegler, A., Henke, T., Wiedrick, J., and Helou, L. B. (2018). Effectiveness of testosterone therapy for masculinizing voice in transgender patients: A meta-analytic review. *Int. J. Transgenderism*, 19(1):25–45.
- Zimman, L. (2012). Voices in transition: Testosterone, transmasculinity, and the gendered voice among female-to-male transgender people. Dissertation, University of Colorado.
- Zimman, L. (2015). Pitch and intonation. In Milani, T. M., editor, *Language and Masculinities: Performances, Intersections, Dislocations*, pages 197–219. Routledge, New York.
- Zimman, L. (2017). Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants. *Linguistics*, 55(5):993–1019.
- Zimman, L. (2018). Transgender voices: Insights on identity, embodiment, and the gender of the voice. *Language and Linguistics Compass*, 12(8).
- Zraick, R. I., Skaggs, S. D., and Montague, J. C. (2000). The effect of task on determination of habitual pitch. *J. Voice*, 14(4):484–489.