Tracking Voice Changes During Testosterone Hormone Replacement Therapy

by

Pam Brown

Submitted in partial fulfilment of the requirements

for the degree of Master of Science

at

Dalhousie University

Halifax, Nova Scotia

February 2025

For my friends, Isaac and Frank, for always inspiring me to be myself.

Table of Contents

List of Tables

List of Figures

Abstract

This thesis aims to describe the acoustic speech changes occurring during the first year of testosterone hormone replacement therapy (T). Using data extracted from sixty YouTube videos posted by transmasculine people documenting voice changes, this study employed a novel recruitment strategy to supplement the concern of low participant numbers in the existing literature. Measures included mean fundamental frequency (MF0) and formant frequencies, used to estimate changes in vocal tract length (VTL).

Results: The MF0 of the transmasculine group reduced by 44%, from an average of 217.47 to 122.81 Hz, in one year. Their resulting MF0 fell within a standard deviation of the mean MF0 of cisgender men. Their VTL increased from an average of 14.14 to 14.90 cm, falling within two standard deviations of the mean of the cisgender men.

This study, containing the largest subject pool to date, adds to the literature on voice changes during T by using publicly available recordings and including estimates of changes to vocal tract length.

List of Abbreviations Used

2SLGBTQI+	Two-Spirit, Lesbian, Gay, Bisexual, Transgender, Queer, Intersex +
AFAB	Assigned female at birth
AMAB	Assigned male at birth
Cis	Cisgender
FF	Formant frequency
F0	Fundamental frequency
F1	First formant
F2	Second formant
F3	Third formant
HRT	Hormone replacement therapy
Hz	Hertz
MF0	Mean fundamental frequency
ST	Semitones
T	Testosterone hormone replacement therapy
Trans	Transgender
VTL	Vocal tract length

Acknowledgements

This project would not have been possible without the support and guidance from my supervisor, Dr. Michael Kiefte. Thank you for your unwavering enthusiasm for this work and for helping and encouraging me through every roadblock along the way.

Thank you to the members of my committee for providing your individual expertise that was instrumental in shaping this project to become what it is. Thanks to Dr. Arati Mokashi, whose expertise in gender affirming care brought a new perspective to this project and enabled me to consider the many factors contributing to outcomes from hormone replacement therapy. Glen Nowell, thank you for sharing your knowledge, ideas, and passion about transgender voice care. Your invaluable experience gained through working with this community has served this project immeasurably. Finally, I would like to extend my deepest gratitude to Dr. Rebecca Affoo for always inspiring me to strive towards academic excellence. From brainstorming methodology, to helping me write two last-minute conference abstract submissions, the impact you had on the success of this project and my growth as a researcher cannot be overstated.

I would also like to thank Shelley McKibbon at the WK Kellogg Health Sciences Library at Dalhousie for your creative problem-solving and guidance with my search strategy on YouTube.

Thank you to my friend, Kevin Shanks, for your help with navigating the technical aspects of this thesis. I appreciate your generosity in talking me through the software confusion. I couldn't have done this without you.

I extend my deepest gratitude to my parents, Tracy and Chris, for believing in me and cheering me on at every personal and academic milestone that led up to the completion of this thesis. I could write another seventy pages just on the appreciation I have for you both, but that would take another three years to complete.

I'd like to extend a huge thank you to my roommates for supporting me throughout this project. Thank you, Antonia, for being a constant source of laughter and joy during some of the more stressful times towards the end of this journey. Thank you, Bria, for always cooking for me when I was neck-deep in writing for days on end. Your feta-tomato pasta was my lifeline some days. Thank you both for picking up my slack with the dishes and cleaning so I could prioritize meeting my deadlines. You have both been such huge pillars of support for me throughout this process, but especially in the recent weeks as I was completing the final steps. I truly do not know how I would have made it to this finish line without you there to cheer me on and manage things for me that I was struggling to manage for myself.

Special thanks to my friends who have helped me reach this landmark. Jenna, thank you for supporting me throughout this degree, for believing in me, and for always being up for discussing theory. Thank you, Capri, for being my first friend in Halifax and making this experience feel like the best TV show.

Finally, I would like to thank my dear friend, Frank, who was with me every step of the way. I am eternally grateful for all that you have done to support me and carry me through the many ups and downs of this thesis. I cannot imagine having spent the past three years on this work without you by my side.

Chapter 1: Introduction

Testosterone hormone replacement therapy, also called ‘T’ by members of the LGBTQ+ community and their healthcare providers, is a key gender-affirming treatment for transgender (trans) men and other transmasculine[1] individuals seeking medical masculinization. Among the many masculinizing effects of T, voice changes have been reported to be of particular importance for trans clients (Hodges-Simeon et al., 2021). Despite the clinical importance of these voice changes, there is not yet a complete understanding of how T impacts speech—either physiologically or acoustically. This is in part due to low participant numbers, and the tendency to focus on fundamental frequency as the primary, or only, measure in relevant studies. For this reason, there is a need for additional research in this area that includes larger participant pools and an investigation into other relevant acoustic properties of speech acoustics such as formant frequencies.

This thesis aims to investigate the effect of T on two acoustic properties of speech, fundamental frequency (F0) and formant frequencies (FF), across a period of one year, as these parameters have been a subject of interest in previous research. This will be achieved using a retrospective study using audio samples obtained from YouTube videos posted in participation with a popular internet trend in the community (often called “This

[1] Transmasculine (or transmasc) is a term used to broadly refer to transgender and/or non-binary people who were assigned female at birth but identify with masculinity (Ferguson, 2021). This term has been selected over “transgender man” to include non-binary participants who take testosterone as part of their transition regardless of their individual gender identities. For more discussion on terminology used in this thesis, please refer to section 1.1.

is my voice on T”). The results will then be compared to reference voice data from cisgender (cis) men and women from a database of continuous speech.

Results from this study will provide the field with more information about the progression of voice changes resulting from T, in turn informing future patients and their healthcare providers about the changes they may expect to see when undergoing treatment.

The research questions this study aims to answer are:

1. How does the voice of a transmasculine individual change in the first year of testosterone hormone replacement therapy in terms of the following measures?
    a. mean fundamental frequency (MF0)
    b. formant frequencies (FF)
    c. vocal tract length (VTL)
2. After one year of T, are there any significant differences in the voices of cis men, cis women, and transmasculine people in terms of the measures listed above?

1.1 Terminology Used in This Thesis

In writing this thesis, the author acknowledges that terminology used to refer to the 2SLGBTQI+ community, especially transgender and other gender diverse people, is in constant flux. It is therefore important to begin with a discussion on terms.

Sex assigned at birth is a term used to refer to the label applied to a person at birth, typically as either male or female. A person’s sex assigned at birth is recorded according to their outward appearance at birth; typically, this is based on the appearance of their genitals, but may also be determined based on chromosomes, gonads, and hormones (Sherbourne Health, 2020). This term replaces the term biological sex, which has been contested by some members of the 2SLGBTQI+ community and medical community as it is unable to capture the diverse biological, anatomical, and chromosomal variations that exist within our species (Planned Parenthood, n.d.; Sun, 2019). Sex assigned at birth also emphasizes that sex is not a simple and fixed identity but can shift and morph with age. For instance, some differences of sex development (previously called intersex conditions) are not identified at birth, and so a person's understanding of their sex can change with new knowledge of their chromosomal and anatomical makeup. Additionally, features of sex, such as hormone levels, gonads, and genitals, can be medically altered, blurring the lines of what determines one's biological sex.

When referring to sex assigned at birth, the terms assigned female at birth (AFAB) and assigned male at birth (AMAB) are used to refer to people categorized as each label. These terms highlight the fact that sex at birth is a label assigned by an outside person, typically a doctor, and may not fully describe one's body as it currently exists.

Transgender is a term used to refer to people whose internal sense of gender is different from their sex assigned at birth. People who are transgender may identify with one or more genders, no gender, or as having a fluctuating gender. The terms used by transgender people are numerous. For simplicity, the term transmasculine is being employed in this paper to refer to a person who was assigned female at birth but identifies with a gender that is different from social expectations of their assigned sex. This may include trans men, non-binary people, genderfluid people, and many more identities. This term was selected for use in this paper based on its popular use within the trans community.

Cisgender is a term used to refer to a person whose gender aligns with societal expectations according to their sex assigned at birth. For instance, a man who was assigned male at birth and a woman who was assigned female at birth would both be considered cisgender.

1.2 Anatomy and Physiology of Speech

Three areas of the body work together to produce speech. These are the respiratory system, the larynx, and the vocal tract. The respiratory system houses the lungs and associated structures which produce air pressure that is forced through the larynx and vocal tract to produce speech. The larynx is composed of many complex structures, including cartilages, tendons, ligaments, and muscles. The vocal folds exist within the larynx and act as the phonatory source for speech. The vocal tract is the hollow space that spans from the outer surface of the lips through the oral cavity and pharynx and ends at the surface of the vocal folds. This is the space through which the sound waves travel from the larynx to the outside world.

According to source-filter theory, speech is composed of two distinct components. The source refers to the structure that vibrates to produce sound waves in the air. In speech, the vocal folds act as the source, and the process of producing sound through vocal fold vibration is called phonation. We produce phonation by forcing air from our respiratory system through the glottis between the vocal folds. As the vocal folds vibrate, they make contact with each other in a rhythmic, periodic pattern at a rate between 70–250 times a second (i.e., 70 – 250 Hz; Fulop, 2011). The frequency at which the vocal folds vibrate is referred to as the fundamental frequency (or, in layperson’s terms, the pitch of the voice). The fundamental frequency is determined by the features of the source: a source with more mass or less tension will produce a fundamental frequency that is lower than a source with less mass or more tension. A speaker can manipulate these features to a degree by contracting or relaxing muscles that stretch or shorten the folds in various dimensions. The extent to which these features can be manipulated depends on many factors, including physiological restraints and neuromuscular control.

The filter refers to a container through which the sound waves produced by the source travel. This container modifies, or filters, the source sound. As the source vibrates, it produces a periodic wave that contains many overtones or harmonics in addition to the fundamental frequency, as the waveform of speech is not a perfect, simple sinusoidal shape such as that of a pure tone. Overtones are always integer multiples of the fundamental frequency—e.g., if the fundamental is 128 Hz, the harmonics or overtones will be 256, 384, 512 Hz, etc. As the overtones are modified by the vocal tract to create sound characteristics unique to an individual, these are influenced by a person's anatomy, language, and social factors. Together they would all influence the timbre of a sound.

In speech, the filter is the vocal tract. This filter can be manipulated to alter the source sound from the larynx in various ways. This is achieved through adjusting the position of the tongue, jaw, velum, and larynx. Depending on the characteristics of the filter, different resonant frequencies are amplified or attenuated. The resonant frequencies that are amplified to possess high levels of energy are called vocal-tract resonances or formant frequencies. The lowest formant frequency present is called F1, the second is F2, the third is F3, and so on. Vocal-tract length is inversely related to the average formant values—a longer vocal-tract will produce lower formants, and vice versa.

In most languages, formant frequencies are what create the primary perceptual difference between vowels. That is, different vowel sounds in a language are characterized by a specific distribution of formant frequencies, and the formant frequencies are what cause us to hear different vowels. For instance, in Standard Canadian English, the vowel /i/ (pronounced as in the word heat) has a low-frequency F1 and high F2; the vowel /ɑ/ (pronounced as in the word hot) has a high F1 and low F2; the vowel /u/ (pronounced as in the word hoot) has a low F1 and low F2; and so on.

Each formant is shaped by different parts of the vocal tract and therefore reflect positional modifications to different areas of the anatomy. Changes in F1 are largely attributable to the configuration of the pharynx which is the region of the vocal tract between the larynx and the velum or soft palate. Therefore, changing F1 is achieved by changing the length of or by constriction of the resonating portion of the pharyngeal cavity. In speech, we can alter this resonating portion by changing the height of our tongue in our mouth which creates a constriction in the pharynx at different heights. Alternatively, although lowering the larynx increases the size of the entire vocal tract, laryngeal lowering disproportionately impacts the first formant, as it more directly increases the length of the resonating portion of the pharyngeal cavity, resulting in a lower F1. Vowels that have a low F1 due to their high tongue position and no constriction in the pharynx are typically called high vowels. Vowels with high F1 and low tongue position are called low vowels. We can also adjust F1 by raising or lowering the position of our larynx which would shorten or lengthen then pharyngeal cavity. Changing the position of the jaw can also contribute to changes in F1; lowering the jaw would effectively lower the tongue's position, shortening the pharyngeal cavity, and raising F1. Raising the jaw would have the opposite effect.

Changes in F2 can be attributed to the configuration of the oral cavity which is bounded by the lips and the velum or soft palate. Therefore, changing F2 is achieved by changing the length of the resonating portion inside the mouth. In speech, this is done by changing the position of our tongue to sit closer to the front or the back of our mouth. Placing our tongue towards the back of our mouth allows for more space within our mouth, increasing the length of the resonating portion of the oral cavity, and produces a lower F2. Placing our tongue towards the front of our mouth has the opposite effect, raising F2. Vowels that have a low F2 due to the tongue's position far back in the mouth are called back vowels. Conversely, high F2 corresponding to vowels with a front position of the tongue are called front vowels. We can also adjust F2 by retracting or protruding our lips, or raising or lowering our jaw, which would either decrease or increase the space in our oral cavity.

Although speakers are well capable of manipulating F3, changes in F3 are much more complex and cannot be associated with a specific region of the vocal tract. Further, speakers have very little control over formant frequencies beyond F3. Although they would otherwise be good correlates to vocal tract length, as they tend to remain static in speech, higher formant frequencies are very difficult to measure due to their low energy and susceptibility to vocal parameters such as breathiness and noise. Therefore, formant-frequency estimates beyond F3 are known to be unreliable.

In summary, the vocal folds are responsible for vocal pitch and the vocal tract is responsible for its timbre. The pitch is what makes a voice sound high or low, and the timbre is what differentiates many speech sounds. Put more plainly, singing the same vowel at many different pitches is done by manipulating your vocal folds, but saying different vowels with a monotone pitch is done by manipulating your vocal tract. The extent to which an individual can manipulate vocal folds and their vocal tract is limited by their individual anatomy and motor control.

Together, characteristics of the source (vocal folds) and the filter (vocal tract) determine the acoustic qualities of a voice. There is great variability across species with respect to the characteristics of these structures, leading to great variability in acoustic output. One major factor contributing to differences between human individuals with regards to speech is sex assigned at birth. During puberty, sexual dimorphism is exaggerated as members of different sexes develop secondary sex characteristics that contribute to perceived differences between them. This is due to differing physiological changes that are produced by differing sex hormones.

Almost all people experience voice changes during puberty; as the body grows, so does the larynx, vocal folds, and vocal tract. However, the extent to which these changes occur depends on sex assigned at birth and the presence of hormones. Cis women and other people assigned female at birth (AFAB) typically undergo an estrogenic puberty. In this population, there are higher levels of estradiol available in the body (Zamponi et al., 2021). Cis men and other people assigned male at birth (AMAB) typically undergo an androgenic puberty, with higher levels of testosterone present. Testosterone is thought to target androgen receptors in the vocal folds, resulting in increased length and thickness of the vocal folds (Zamponi et al., 2021). As AMAB people have higher levels of testosterone, this process occurs to a greater extent in this population. Changes in the vocal tract are also greater in AMAB people, as testosterone is also believed to be responsible for the so-called secondary descent of the larynx during androgenic puberty, resulting in increased vocal tract length and lower formant frequencies (particularly F1) for AMAB people (Markova et al., 2016; Zamponi et al., 2021). In general, the thicker vocal folds and longer vocal tracts found in AMAB people results in lower fundamental frequencies and formant frequencies than in AFAB people.

1.3 Voice and Gender Perception

The available data for fundamental frequency for cis men and cis women have led to differing normative estimates over time. Depending on the study, average mean fundamental frequency has been reported to be somewhere between 100 and 131 Hz for young adult cis men and between 205 and 220 Hz for young adult cis women (Fitch & Holbrook, 1970; Gelfer & Bennett, 2013; Goy et al., 2013; Hillenbrand & Clark, 2009; Skuk & Schweinberger, 2014).

Goldstein's study in 1980 is a frequently cited work when discussing normative formant frequency and vocal tract length values for adult cis men and women (1980). This study states that the average vocal tract length is 16.9 cm for an adult cis man and 14.1 cm for an adult cis woman. Using this average vocal tract length estimate, we can calculate average F1, F2 and F3 values for cis men to be approximately 509, 1530, and 2540 Hz respectively. Average F1, F2 and F3 values for cis women sit at approximately 610, 1830, and 3050 Hz respectively. This means that the formant values for individual vowels would vary around these averages. The neutral vowel, or schwa (/ə/), would theoretically be produced with F1, F2, and F3 values at approximately these averages.

As a result of sexual dimorphism of the speech mechanism, speech acoustics plays a very significant role in gender perception. While the majority of the literature focuses on fundamental frequency's role in gender perception (Leung et al., 2018), it has been argued by some that formant frequencies play as important, if not more important, a role.

In a study by Gelfer and Bennett (2013), changing a speaker's fundamental frequency was insufficient for changing their perceived gender: cis women were identified as women even when their fundamental frequency was shifted to be within the typical range of a cis man. Conversely, the contribution of formant frequencies to perceived gender was especially important when a speaker's fundamental frequency was modified to be within a gender-ambiguous range of 145–165 Hz. When a person's perceived gender was not judged consistently by listeners, nearly two thirds of the variance in gender perception amongst listeners could be attributed to formant frequencies. Compared to 41.6% of variance in listener perception being explained by fundamental frequency (Leung et al., 2018), this notion points to formant frequencies as being more predictive of perceived gender than fundamental frequency.

However, the literature investigating the relative effectiveness of altering either fundamental frequency or formant frequencies on perceived speaker gender has been quite mixed. A study by Skuk and Schweinberger (2014) found that fundamental frequency had a stronger predictive power for perceived gender than formant frequencies. Still, the same study found that modifying the sound spectrum (i.e., loudness of specific frequencies) and formant frequencies together is sufficient for changing listener's perception of gender even without any change in fundamental frequencies.

Despite the lack of consensus on the relative importance of fundamental frequency and formant frequencies on perceived gender, the literature does support the notion that a combined change in both acoustic properties is more effective in changing gender perception than a change in just one property on its own (Leung et al., 2018).

A study by Merritt and Bent (2024), which compared the relative impact of fundamental frequency, formant frequencies, and articulatory patterns for both cis and trans speakers, found that shifting the fundamental frequency and formant frequencies of trans speakers into the cis reference ranges of their gender (i.e., shifting a transmasculine person's voice to fall within the range of a cis man's and shifting a transfeminine person's voice to fall within the range of a cis woman's) was effective in changing their perceived gender. Interestingly, this study found that when the same changes were applied to cis speakers' voices, there was less consistency in their perceived gender. For cis men, they were substantially less likely to be continued to be rated as men with these parameters shifted, but for cis women, the shifted voices resulted in less of a change from baseline gender judgements. The authors attribute this mixed result to differences in articulation and intonation patterns between the cis and trans groups; the trans groups likely had already adapted speech patterns that match their gender, and so the shift in fundamental and formant frequencies was more effective. In contrast, the cis women had such distinct intonation patterns that shifting the acoustic features of their voices alone was not enough to create a change.

1.4 Testosterone Hormone Replacement Therapy and Voice

While testosterone is thought to produce the sexually dimorphic changes in voice during androgenic puberty in adolescence, it is not yet fully understood how these changes differ in the trans population. Previous research has investigated the way these changes occur in transmasculine people receiving testosterone hormone replacement therapy. Following a thorough search for relevant literature and review of a meta-analytic review by Ziegler and colleagues (2018), eight studies that involved tracking these changes were identified.

A study from Van Borsel et al. (2000) provided the first longitudinal exploration into the changes induced by T in this population, laying the foundation for further research on this topic. The study consisted of two participants (22 and 37 years old) whose voices were recorded once every two months for a period of 17 and 13 months, respectively. This study found that MF0 significantly reduced, though by less than one octave (a doubling or halving of a frequency) each, with the most significant changes for both subjects occurring at approximately 4 months after treatment onset and smaller changes continuing throughout the study.

Nearly a decade after Van Borsel et al. (2000), Damrose (2009) published a longitudinal case study of a 33-year-old semi-professional jazz singer. The singer's MF0, extracted from sustained /ɑ/, was collected prior to the start of treatment, then monthly for the first 6 months, then every 3–6 months for a combined total of 16 months of treatment. Mirroring results from Van Borsel et al. (2000), Damrose's study reported MF0 significantly decreased after 16 months, this time by approximately 1 octave from 228 Hz to 113 Hz (SD ±27.46), with the largest change occurring at around 4 months and then stabilizing.

In an attempt to address the absence of prospective longitudinal studies on this topic with more than just a couple subjects each, Deuster et al. (2016a) published a study conducted in Germany with the largest subject pool at that time. It was also the first study of its kind to include a control group of cis men for comparison. The study started off with 15 participants but, due to attrition, ended with nine (20–33 years old). The control group consisted of 21 cis men (20–81 years old). Each participant was asked to perform self-recordings at home once a week. Results of F0 were reported in Hz and in semitones (ST). They found that F0 was significantly reduced and was not significantly different to that of the cisgender control group (starting median F0: 192 Hz; ending median F0: 108 Hz) and the largest changes in F0 occurred between weeks 8 and 12, followed by weeks 12 and 16. In terms of change in F0 in ST (mean: 8.93), they note that their results supported findings of Van Borsel et al. (2000) who reported that the total change in 1 year is less than an octave; they suggest that Damrose's case study (2009) of the jazz singer could have shown larger changes due to the subject's assumed elevated level of voice control. By recruiting a larger group—including a control group—and conducting more frequent recordings, this study was considerably stronger than the prior two studies of its kind.

Adding to the literature on the correlation between F0 and voice satisfaction, Nygren et al. (2016) published a longitudinal study conducted in Sweden. This study is particularly notable due to its large number of participants and its span of 24 months, both made possible because of the retrospective nature of the study. Medical records from the speech-language-pathology department of the local hospital were accessed. Out of 104 identified patients, 50 (18–64 years old) had sufficient recordings to be included in the study. Recordings were selected and grouped into clusters that represented approximately 3, 6, 12, 18 and 24 months on T. Acoustic measures included F0 (mean, mode, minimum, maximum). As expected, results showed a significant effect of time on MF0, F0 mode, and minimum F0, all of which saw large changes over the course of 0–3 months and 3–6 months. Group averages for MF0 and F0 mode continued to decrease for the remainder of the year and then stabilized with no changes occurring after 12 months. There was a moderate-to-good relationship between voice satisfaction ratings and MF0. The primary advantage of the Nygren et al. study is the large number of participants; it is currently the largest study of its kind.

The next study was published by Irwig et al. (2017): a prospective longitudinal study with the goal of understanding the timing and degree of MF0 changes occurring over 12 months in an ethnically heterogeneous sample. Their study had seven trans men aged 18–39 years from three different racial backgrounds: black, white, and mixed-race. Voice assessments consisting of three sustained /ɑ/ vowels and a reading of The Rainbow Passage (Fairbanks, 1960) were collected in the lab twice at baseline, and once each at 3, 6, 9, and 12 months in treatment. The only measure studied was MF0, which was measured in Hz and in ST, as suggested in Deuster et al. (2016). The authors found that, at baseline, four participants had MF0 in the typical cis woman's range and three had MF0 in the gender neutral/high cis man's range (Brown et al., 1993; Colton et al., 2011; Howard & Murphy, 2008; Morris et al., 1995, as cited in Irwig et al., 2017). The mean decrease from 0–3 months was 2.2 ST (19 Hz), 0–6 months was 4.7 ST (37 Hz), 0–9 months was 5.8 ST (44 Hz), and 0–12 months was 6.4 ST (49 Hz). The variability for each of these means was quite large. At 12 months, for instance, decreases in MF0 ranged from 2–10 ST or 17–79 Hz. At 6 months, all seven participants experienced significant decreases in MF0, and by the end of 12 months they were all within the cis male range after experiencing decreases that were, again, less than 1 octave. This means that smaller changes in MF0 for some participants were simply due to possessing deeper than typical voices at baseline. They found no differences based on race/ethnicity, and they note that baseline MF0 was not predictive of MF0 after 12 months.

They were able to identify four patterns of voice changes: minimal MF0 changes from months 0–3, MF0 changes from 0–6 months only, MF0 changes from 3–9 months only, and consistent MF0 changes through all 12 months. This indicates that there is notable variation in the timeline over which someone's voice undergoes changes—something that had not been noted in prior studies—though they note that further studies with larger sample sizes should be conducted to investigate which pattern is most common. This study provided a new perspective on the timeline of MF0 reductions in this population, challenging the previous understanding that 2–5 months is where the largest changes occur. However, due to its small sample size and narrow focus, further investigation is required.

Hancock et al. (2017) published a study that investigated the relationship between self-perception and changes in voice properties of the same seven participants from Irwig et al. (2017). Acoustic measures included MF0 in Hz and habitual pitch in ST. The recordings involved sustained /ɑ/, three glissandos up to the highest phonation frequency, and three down to measure vocal range, and reading The Rainbow Passage (Fairbanks, 1960).

The results of Hancock et al. (2017) reflect the pattern seen in the studies prior: there is high variability between individuals. While all participants self-reported a neutral or female sounding voice at baseline, they each perceived their voice to have become more masculine at the end of the study. Self-perceived masculinization started at just 3 months at which point every participant self-reported a neutral or male-sounding voice, and four participants continued experiencing self-perceived masculinizing effects after that point. Self-perception measures did not correlate with any acoustic measure at any point during the study, a finding that conflicts with Deuster et al. (2016b).

The next study of note was conducted by Cler et al. (2020). This longitudinal case study of a 30-year-old trans man in his first year of T was the first study to include formant frequencies in its analysis. Baseline data was collected three times, and then data was collected every 2 weeks for a year. Collection tasks included isolated vowels (/ɑ/, /i/, /u/, /æ/, /e/, /o/), spontaneous speech, several standardized scripts, such as the Rainbow Passage (Fairbanks, 1960) and CAPE-V sentences (Zraick et al., 2011), and singing the vowel /i/. From these tasks, they analyzed more measures than did any other study: F0 (mean, minimum, and maximum, in Hz and ST), habitual F0 (as in spontaneous speech), formant frequencies, and estimated laryngeal height, among several others. Additionally, they included subjective ratings of voice quality from an expert and gender perception ratings which were made by a group of eight adult laypeople.

Results of Cler et al. (2020) showed significant changes in MF0 while reading, decreasing from 183 Hz to 134 Hz—a 5.4 ST difference. Changes in habitual F0 was even greater: at baseline, 211 Hz, reducing to 137 Hz—a difference of 7.5 ST. The largest changes for this participant occurred between months 1–2 and 3–4; and months 3–4 and 5–6. The most interesting and novel measure in this study—formant frequencies—decreased significantly (the fourth formant changed from 3693 to 3554 Hz), demonstrating an increase in estimated vocal tract length of 0.64 cm (16.26 cm at baseline to 16.90 cm at end). This is a very important finding that suggests that T does more than just thicken vocal folds to affect F0, but also has the potential to reduce formant frequencies by lengthening the vocal tract as seen in cis male puberty. The final length of 16.90 cm is also within norms for cis men (Goldstein, 1980, as cited in Cler et al., 2020). In terms of gender perception ratings, the participant was reliably identified by a group of naïve listeners as female until 15 weeks on T, and then reliably as male after 37 weeks, despite ending with a MF0 that is still considered slightly higher than the average for cis males as noted in previous studies as 123 Hz. This study, the first to include formant frequencies, points to a real need for further investigation in this area with a larger sample size to see how formants may be impacted by T throughout this population.

The final study discussed here was conducted by Hodges-Simeon et al. (2021). This study aimed to determine the importance of changes in vocal masculinity for trans men, as well as the effectiveness of T in masculinizing voice. This study was the first to include an analysis of formants with a larger group: 30 trans participants aged 20–40 years old took part in the study, four of which identified as non-binary and the remaining 26 as men. Each trans participant had been taking T for at least 9 months at the start of the study, making this the only cross-sectional study in this review. Two control groups were used, one with 122 cis men aged 21–28 years old and the other with 32 cis women aged 21–33 years old. Each trans participant completed a survey that touched on voice masculinity, body fat distribution, body hair, facial masculinity, muscle mass, strength, sex drive, self-identity, and psychological masculinity. The survey asked about their satisfaction with these traits pre-T, the importance of seeing changes in them while taking

T, and the amount of change they have observed in these traits since starting T. Voice recordings, conducted once each by members of all groups, consisted of a recitation of The Rainbow Passage (Fairbanks, 1960), counting from one to ten, and producing a sequence of vowels (/ɛ/, /i/, /ɒ/, /o/, /u/). Recordings were analyzed to measure MF0, F0 variance, and VTL estimations using the first four formants of just /ɛ/ and /ɒ/.

Results of Hodges-Simeon et al. (2021) found that voice masculinity was among the lowest scoring traits in terms of satisfaction pre-T and the highest scoring trait in terms of importance. As expected, they also showed that T is effective at masculinizing vocal pitch, with MF0 and F0 variance in the trans group being significantly lower than the cis women control group and not significantly different from the cis men. In terms of formants, the trans group had significantly longer vocal tracts than the cis women, but still shorter than the cis men, suggesting that, on average, the changes occurring from T were not as great as the changes that had occurred during androgenic puberty for the cis men. However, this study had a minimum inclusion criterion of just 9 months of T compared to the 12 months covered by the case study from Cler et al. (2020). As that case study reported, formant frequency changes were occurring throughout the entire 12 months, so it is possible that the smaller vocal tract changes reported by Hodges-Simeon et al. could be attributed to the shorter length of therapy. This study provides support to the notion that formant frequencies are, in fact, significantly lowered as a result of T; however, the extent to which this change occurs after 9 months is still unknown.

In summary, these eight studies have contributed considerably to the research by providing estimates of acoustic changes in speech to be expected after a year of testosterone therapy. From the first study in this body of literature (Van Borsel et al.,

2000) to the most recent (Hodges-Simeon et al., 2021), the most consistent finding is that T is highly effective at lowering the vocal pitch of transmasculine people, as evidenced in the change of MF0 from baseline to end. There was large variance across studies, however. The largest change was found in Damrose's (2009) case study of one 33-year-old semi-professional singer whose MF0 dropped 115 Hz (from 228 Hz at baseline to 113 Hz at 16 months). Contrastingly, one 37-year-old participant in the study from Van Borsel et al. showed a much more conservative change of just 29 Hz (from 181 Hz at baseline to 152 Hz at 13 months). This variance could be attributed to a few factors, including the differences in baselines, the voice training of the singer allowing for more vocal control, and the effect of age on the larynx's ability to change. Nevertheless, while a reduction of 29 Hz is small in comparison to results from other studies, it is still a significant change. The remaining studies showed MF0 reductions ranging from 49 to 84 Hz with post-treatment group mean MF0s ranging from 108 to 152 Hz (Cler et al., 2020; Deuster, Matulat, et al., 2016; Hancock et al., 2017; Hodges-Simeon et al., 2021; Irwig et al., 2017; Nygren et al., 2016; Van Borsel et al., 2000). These findings support the conclusion that T results in significantly lower voices than baseline. Additionally, each study reported final group mean MF0 values within the typical cis male range of 85–155 Hz as reported in Fitch & Holbrook (1970, as cited in Cler et al., 2020; it is mentioned in Cler et al.'s paper that these reference values are disputed and may be too broad).

Finally, with just two studies looking at formant frequencies, a pattern emerges suggesting that VTL does increase as a result of T, though the question remains whether the extent of this increase meets cis male reference values, as results have been conflicting thus far (Cler et al., 2020; Hodges-Simeon et al., 2021).

Previous literature has generally relied on small sample sizes, partially due to drop-out rates caused by the demand of participating in longitudinal studies, and partially due to the narrow criteria for appropriate subjects, as all but one study required participants to be naive to T at the start of data collection in order to obtain a baseline. When the sample size is larger, studies have examined a more restricted collection of measures, the most common being MF0. For future studies, a focus on larger sample sizes, the use of control groups, and the inclusion of more acoustic parameters such as formant frequencies, is necessary to fill the gaps in the literature.

1.5 "This is My Voice on T": Trans Voice on YouTube

As the second most-visited website on the Internet, YouTube is the most popular video-sharing platform currently operating online (Top Websites Ranking, 2025). Since as early as 2006, YouTube has been used by trans vloggers as a platform to post content about their lives (Raun, 2015). One common video format used to document voice changes that take place following the start of HRT is sometimes referred to as "This is my voice on T" or "This is my voice on testosterone." In this format, trans people document their voice transitions online by posting videos of themselves speaking at different times following the start of treatment. It is standard to begin each video clip with a verbal statement of the amount of time since the commencement of HRT; "This is my voice pre-T/testosterone", "This is my voice one day on T", "This is my voice five weeks on T", etc. The frequency of these clips is highly variable both between and within individuals. Some choose to record themselves on a weekly basis, others monthly or annually. Some record at inconsistent intervals, most typically recording with more frequently early in treatment and then less frequently as changes stabilize over time.

To date, acoustic data from these videos have been used just once in previous literature. The study extracted acoustic data (MF0, F0 contour, and formant frequencies) from two speakers (Groll et al., 2022). This data was used in speech resynthesis in a study investigating listener perception of the gender of transmasculine voices.

Aside from this study, no studies have yet explored the use of the abundance of data available on YouTube to address the issue of low participant numbers that is frequently noted in research involving this population.

Chapter 2: Methods

2.1 Study Design and Participants

The study described here is a retrospective quasi-experimental design. The population of interest consists of trans YouTubers whose acoustic data was extracted from publicly available YouTube videos. The data from the trans group was compared to data from two control groups which consisted of cis men and women, and further compared to reference data that was acquired from a large database of speech from cis men and women.

This study aimed to collect a large number of self-published YouTube videos—up to 100 unique speakers. A strict search procedure to identify relevant videos for inclusion in the study was employed by the researcher on February 6, 2024. Details of this procedure are described in the following section.

2.1.1 Search Procedure

The search procedure consisted of thirty-six searches on YouTube using twelve different search terms. For each term, three searches were performed, with each of the three searches' results ordered in one of the possible ways: by view count, by date, and by relevance. Ordering the videos in three different ways aimed to reduce the impact of popularity on the results of the search and capture as many unique results as possible. Videos were then identified and saved if, after a quick review of the video contents, they appeared relevant according to the screening criteria. Each search ended once results showed five irrelevant videos in a row. Videos were then reviewed for duplicates and duplicates were removed.

2.1.2 Search Terms

The twelve search terms used are listed below. The first four are terms that were used in a sociological study by Borck and Moore (2019). The remaining eight were developed in collaboration with a librarian at Dalhousie University after preliminary search attempts to identify which terms resulted in the most relevant videos. All searches were conducted using an incognito window on Google Chrome and an anonymized Google account to remove the influence of cookies and account data, as suggested by Sui et al. (2022).

The twelve search terms were:

1. "transgender testosterone"
2. "trans man testosterone"
3. "FTM testosterone"
4. "trans t shot"
5. "this is my voice on T"
6. "testosterone voice comparison"
7. "transition voice comparison"
8. "trans voice comparison"
9. "HRT voice comparison"
10. "trans FTM comparison"
11. "trans FTM HRT"
12. "FTM voice"

2.1.3 Inclusion and Exclusion Criteria

Videos were included in the study if they (a) document a voice comparison consisting of at least two dates spanning at least six months and (b) contain a mention of the time on testosterone for each recording. Exclusion criteria rejected videos that (a) contain excessive background noise (e.g., people talking in the room, loud wind or traffic sounds) or music that would interfere with acoustic analysis, (b) are duplicate videos, or (c) are posted by the same YouTuber that posted a video that was already identified. In the case of the same creator posting two videos, the video documenting the longer treatment time was kept.

2.1.4 Control Groups and Reference Data

It was planned to recruit two control groups of 25 cis men and 25 cis women between the ages of 18 and 35 from the Dalhousie community through posters and word of mouth. The age cut-off for the control group was selected as it appeared to best represent the speakers in the videos, and voice acoustics is impacted by age. The recruitment strategy was selected due to convenience of location and scheduling for sessions on site at the university. Interested participants were asked to complete a self-screening questionnaire asking about their experience in puberty, their voice, and related medical history. They were also asked to complete the Voice Handicap Index (VHI) questionnaire (Jacobson et al., 1997) and self-screen if they scored a 2 or above (i.e., responded “yes” to two or more statements). Inclusion criteria selected those who: (a) are cisgender, (b) do not have any concerns about their voice that would indicate a voice disorder (based on their responses on the VHI), (c) have not received any formal voice therapy or training, and (d) have no history of taking supplemental testosterone or several additional medications that are known to impact androgen levels. A detailed copy of the screening questions can be found in Appendix A.

Unfortunately, recruiting sufficient control group participants was not feasible due to time restraints. Therefore, acoustic data extracted from a database of continuous speech was used for comparison. This database is the TIMIT Acoustic-Phonetic Continuous Speech Corpus (Garofolo et al., 1993).

The TIMIT database, named after its sponsor, Texas Instruments, Inc. (TI), and the research institution, the Massachusetts Institute of Technology (MIT), is a corpus of continuous speech from 630 speakers across eight dialect regions. Each speaker in this corpus has been recorded reading ten sentences designed to capture all phonemes in the English language. The "test" portion of the corpus was used in this study.

2.2 Procedures and Measures

2.2.1 YouTube videos and Recordings

Once eligible videos were identified, the audio from each video was downloaded as a WAV file using the opensource software YOUTUBE-DL (Amine et al., 2021). From there, audio files were uploaded to PRAAT V 6.3.20 (Boersma & Weenick, 2023). Pauses were cut from the file to prevent any background noise from influencing the frequency analysis. Each file was then manually annotated using PRAAT's TextGrid feature. This feature allows users to divide an audio file into individual segments which can be then labeled with text. Each file was segmented according to days on testosterone and given a number label (i.e., when the person in the video stated, "this is my voice before testosterone", this became its own segment and was labeled as "0", "this is my voice one month on testosterone" was segmented and labeled as "30", and so on). For the sake of consistency when labeling, one month was counted as 30 days and one year was counted as 365 days.

Members of the two control groups were recorded in a soundproof booth using a Marantz PMD671 professional solid-state recorder and a Shure SM10 headset microphone with the microphone placed approximately 2 cm from the corner of the speaker's mouth. Each participant was asked to speak the sentences "This is my voice on testosterone" and "This is my voice on T" for three trials each. These sentences were selected as they best matched the most common phrase spoken in the YouTube videos. The second trial of each sentence was used for analysis. Participants were also asked to speak a portion of The Rainbow Passage (Fairbanks, 1960). The specific segment of the passage can be found in Appendix B.

Audio recordings were uploaded to PRAAT and annotated using the TextGrid feature to label each segment of the recording according to the stimuli spoken. Trials one and three from each sentence were cut from the file.

Custom PRAAT scripts were used to extract the relevant acoustic information from each file. These scripts can be found in Appendix C, adapted from The Speech Corpus Toolkit for Praat (SpeCT) (Lennes, 2003). For the YouTube videos, the acoustic data was extracted from each segment. For the control groups and TIMIT reference data, the scripts extracted the value from the whole recording because the cis speakers were each only recorded once.

2.2.3 Acoustic Analysis

2.2.3.1 Mean Fundamental Frequency

The script to extract fundamental frequency used the To Pitch function in PRAAT. The timestep was set to 0.01, pitch floor was set to 75 Hz, and pitch ceiling was set to

400 Hz. This script extracted the mean fundamental frequency from the full duration of each segment according to the boundaries set with the TextGrid feature (i.e., mean fundamental frequency was extracted from the entire segment that was labeled “0”, the entire segment labeled “30”, and so on). For the control groups and reference data from TIMIT, all the files for each individual speaker were concatenated and the measurements were taken from the concatenated file.

2.2.3.2 Formant Frequencies

The script to extract formant frequencies used the To formant (burg) function in PRAAT. The timestep was set to 0.01, maximum number of formants was set to 5, window length was set to 0.025, and preemphasis from was set to 50.

For the YouTube videos, the formant ceiling was set to 5500 Hz for all time periods. The formant ceiling determines the range within which the algorithm will detect formants. The PRAAT manual emphasizes the importance of setting the formant ceiling to a value suitable to the speaker, suggesting using a formant ceiling of 5500 Hz for adult women and 5000 Hz for adult men. This is because selecting too narrow or too wide of a range will impact where the algorithm determines whether a formant exists. For instance, setting the formant ceiling too high might cause the algorithm to ignore lower formants and produce artificially high results. This is especially true with back vowels where F1 and F2 are close together. Setting the formant ceiling too low might cause the opposite effect. For the sake of consistency, the value of 5500 Hz was selected for the formant ceiling. While this may have the effect of artificially raising the formant frequency estimates, especially if the vocal tract length increases the speakers’ formant frequencies lower, this more conservative approach reduces the risk of falsely reporting larger changes. This is discussed further as a limitation in the discussion section.

For the control groups and TIMIT reference data, the formant ceiling was set according to the value that matched the assigned sex of the speaker: 5500 Hz for the cis women and 5000 Hz for the cis men.

The script extracted median values for the first three formants (F1, F2, F3). Median values were selected to minimize the influence of extreme outliers.

The formant values extracted from recordings were used to calculate VTL.

2.2.3.3 Vocal Tract Length

Vocal tract length estimates were calculated using the equation for the length of a tube closed on one end, found in Equation 1 below, where n is the formant number, F_n is the frequency of the formant in Hz, and c is the speed of sound in air (approximately 344m/s). An estimated VTL was calculated from each formant separately and then an average of the three estimates was calculated and used as the final VTL.

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

Equation 1 Formula for estimating vocal tract length

2.3 Statistical Analysis

Linear regressions with a log(day) transformation were calculated for each YouTube video for both MF0 and VTL. A repeated measures t-test was conducted on the group's regression coefficients to analyze deviations from zero indicating "no change". An average regression coefficient was calculated across speakers to find the average change over time for the trans group.

Chapter 3: Results

3.1 Participants

3.1.1 YouTube Videos

A total of 142 videos were identified using the search procedure. After removing those that did not meet criteria, 60 videos were included in the study. Figure 1 shows a flowchart of the search results and video selection.

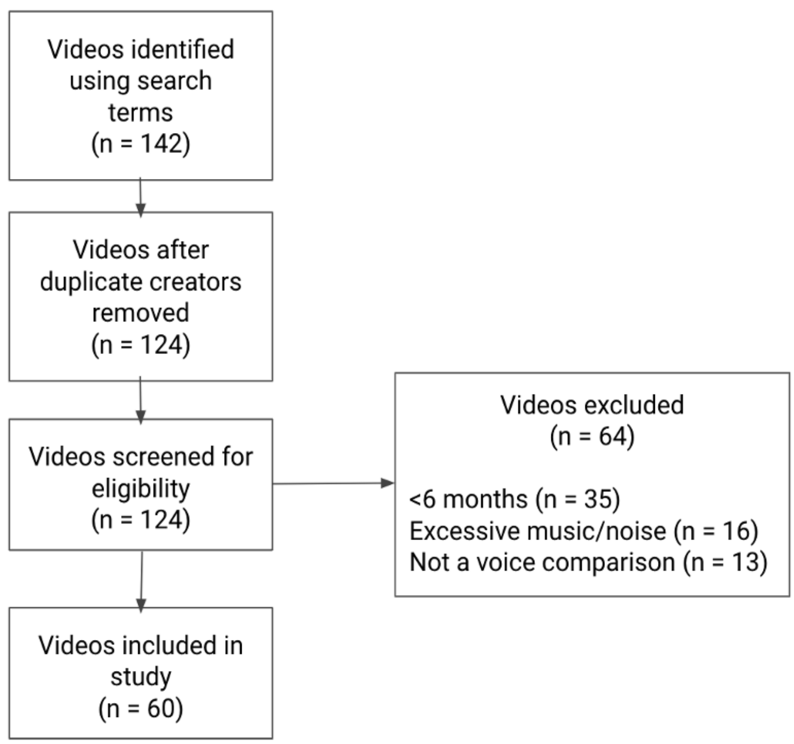

Figure 1 Video selection flowchart

With this study's inclusion criteria stating that eligible videos must document at least 6 months, timespans of included videos ranged from 6 to 114 months on T. The most common length of documented change was 12 months (14 videos), followed closely behind by 6 months (13 videos). See Figure 2 for the distribution of videos according to total length of documented change.

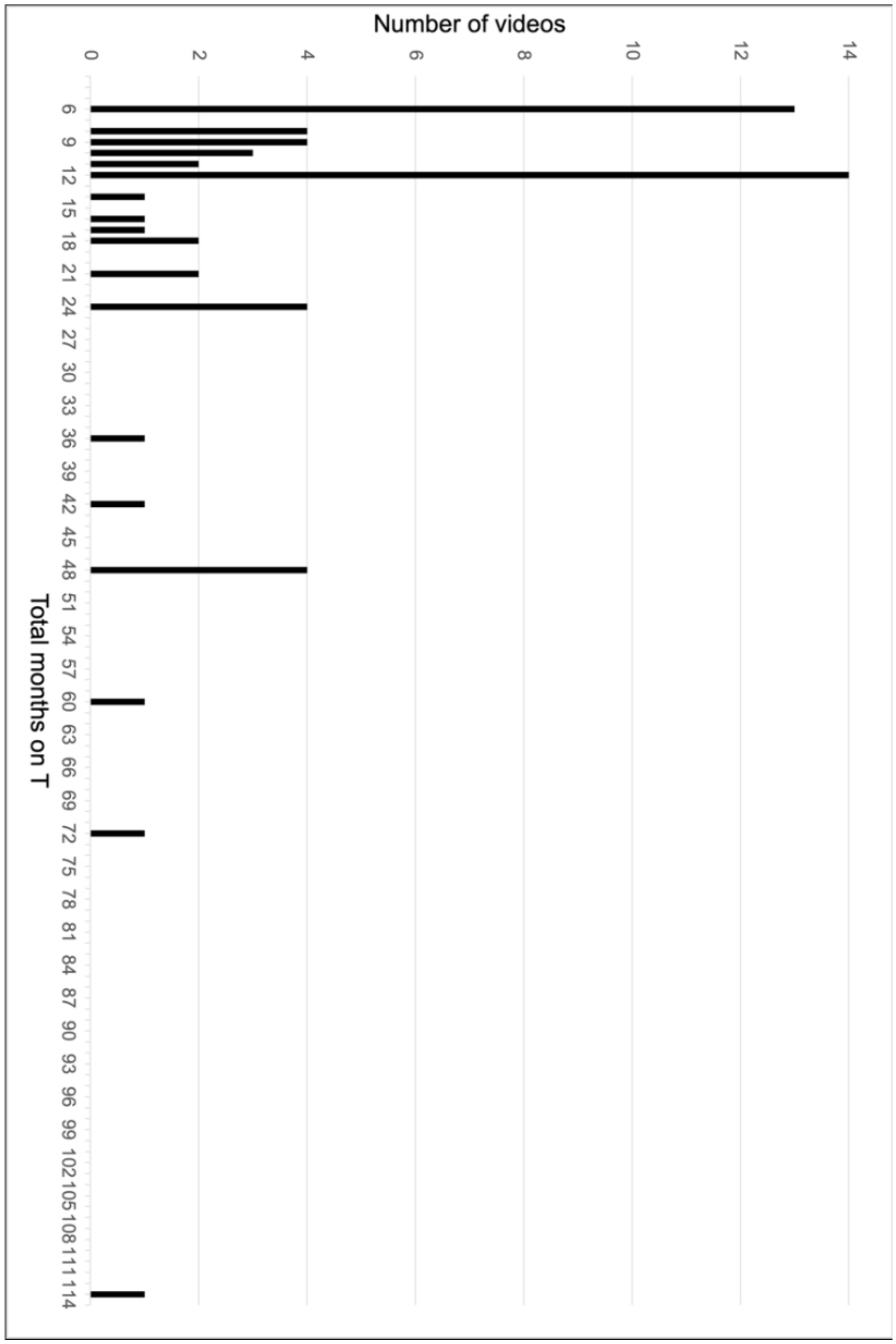

Figure 2 Number of videos according to total number of months on T

The recording schedule used in each YouTube video was highly variable. The most common schedule was monthly recordings, which can be seen by noting the location of the peaks in the histogram in Figure 3.

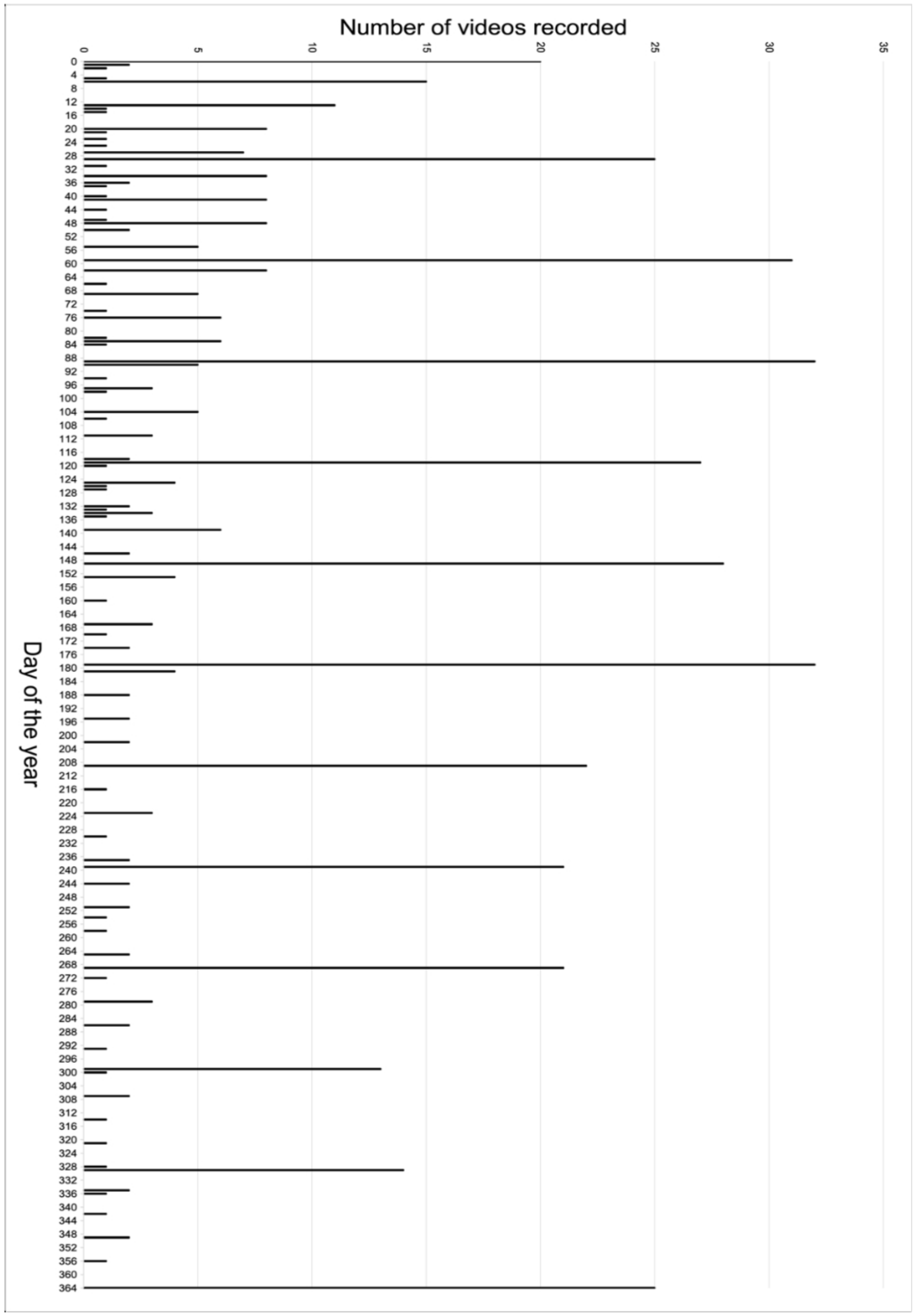

Figure 3 Histogram of clips recorded according to days on T. Each line depicts the number of YouTube videos that contained a clip that was recorded on a specific day during over the year. The largest peaks all correspond to months (i.e., day 0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 365).

After binning by month according to the conversion in Table 1, Figure 4 shows how many videos were posted within ±15 days from a given month.

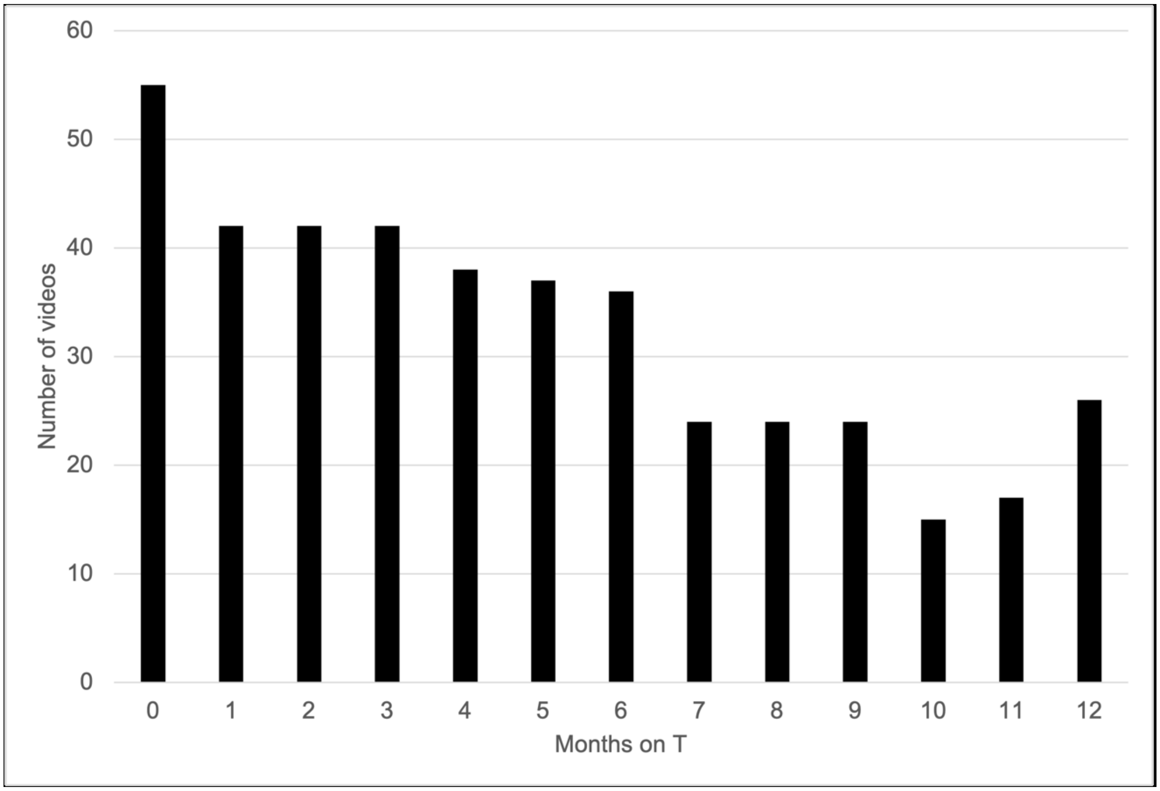

Figure 4 Histogram of clips recorded according to months on T. Each line depicts the number of YouTube videos that contained a clip that fell within ±15 days from a specific month.

Month	Day Range	Month	Day Range
0	0–14	7	195–225
1	15–45	8	225–255
2	45–75	9	255–285
3	75–105	10	285–315
4	105–135	11	315–345
5	135–165	12	350–380
6	165–195	—	—

Table 1 Binning conversion convention

Due to the variable recording schedules in the YouTube videos, longer treatment duration did not necessarily correspond to a larger number of clips in a video (Figure 5).

In fact, the video documenting the longest change (114 months) contained just two clips. The video with the greatest number of clips (53) documented a change of 12 months.

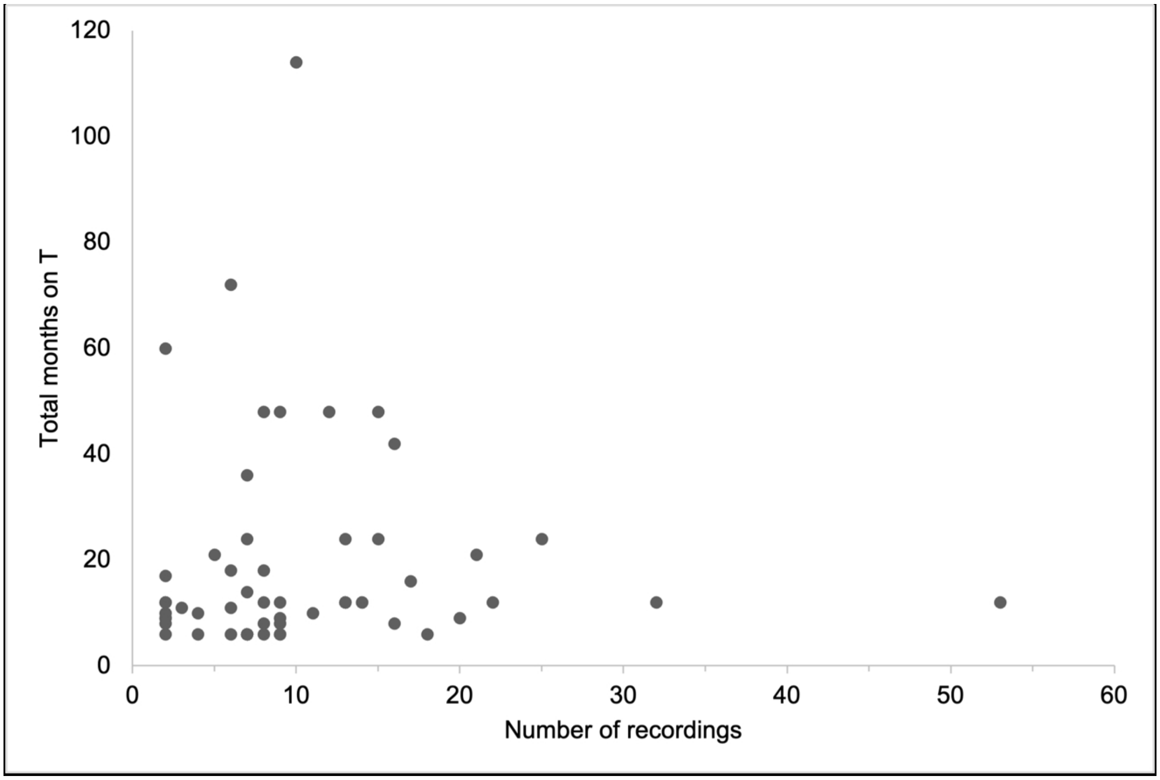

Figure 5 Number of recordings vs total number of months documented

3.1.2 Control groups

Twenty-eight participants (11 cis men and 17 cis women) were successfully recruited to take part in the control groups. Details of the age distribution of the two control groups can be seen in Table 2 below.

	Age (years)		
	Mean	Median	Range
Cis men (n = 11)	25.09	25	24 – 27
Cis women (n = 17)	25.65	26	21 – 30

Table 2 Age distribution of the control group speakers

Due to the inability to recruit sufficient participants in the control groups, results from these groups are being reported but not used to compare against the trans group.

3.1.3 Speech Database

The TIMIT test database (Garofolo et al., 1993) provided data for 112 cis men and 56 cis women. Demographic details, including age and height, of the TIMIT test database can be seen in Table 3 below.

	Age (years)			**Height**		
	Mean	Median	Range	Mean	Median	Range
Cis men (n = 112)	30.36	27	22 – 64	5’11”	5’11”	5’4” – 6’8”
Cis women (n = 56)	30.14	26.5	21 – 67	5’6”	5’6”	5’0” – 5’11”

Table 3 Demographic details of the TIMIT test database speakers

3.2 Acoustic Measures

3.2.1 Fundamental Frequency

Prior to the start of T, the transmasculine people from the YouTube videos had an average MF0 of 217.47 Hz. According to the results of the regression analysis, MF0 was reduced on average by 16.04 (SD: 8.04) Hz per log(day). This is highly significant at $p < .001$. This amounts to an average reduction in MF0 of 94.66 Hz over the course of one year or nearly 10 semitones, or approximately 5/6 of an octave. This means that, on average, the voice will drop in pitch by nearly half of its original fundamental frequency after one year (annual change of –44%). Detailed results from this regression analysis can be found in Table 4. The results for MF0 for all groups can be found in Table 5. A normal distribution curve of MF0 from the trans group and the cis men and women from the TIMIT data is shown in Figure 6.

	Mean Fundamental Frequency
Mean β (SD)	–16.04 (8.04) Hz/log(day)
Standard error	1.04
t	–15.46
p	< .001 ***
Mean at day 0	217.47 Hz
Mean at day 365	122.81 Hz
Annual change	–94.66 Hz (–9.89 ST)
% Annual change	–44%

Table 4 Regression analysis results for fundamental frequency. SD: standard deviation. *** indicates statistical significance at $p < 0.001$.

		Mean Fundamental Frequency (Hz)		
		Mean (SD)	Median	Range
Trans group	Transmasc people (n = 60)	122.81 (25.68)	119.48	81.12 – 202.90
Control groups	Cis men (n = 11)	111.95 (15.00)	109.02	91.20 – 140.79
	Cis women (n = 17)	190.70 (20.65)	197.75	147.74 – 215.22
TIMIT database	Cis men (n = 112)	119.36 (16.25)	118.15	85.05 – 177.78
	Cis women (n = 56)	200.76 (24.12)	198.72	143.12 – 247.53

Table 5 Results for mean fundamental frequency. The results reported for the trans group refer to group values after one year. SD: standard deviation.

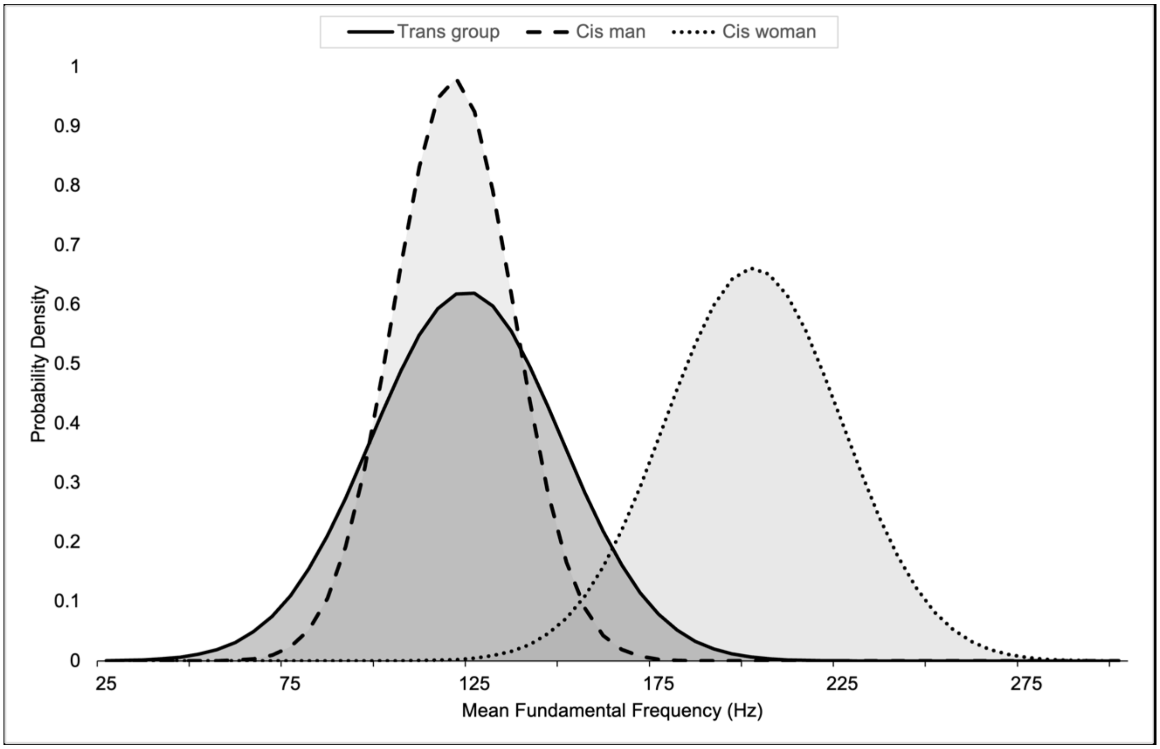

Figure 6 Distribution of mean fundamental frequency

3.2.2 Vocal Tract Length

Statistically significant changes were detected in nearly all VTL estimates. When VTL was calculated as an average from F1, F2, and F3, average pre-T VTL was found to be 14.14 cm. It increased by 0.12 cm (SD: 0.25) per log(day), resulting in an average VTL of 14.90 cm after one year. This was highly statistically significant ($p < 0.001$). This amounts to an annual increase of 0.76 cm, or 5.4%.

After confirming T did result in increased VTL, VTL was then estimated from individual formant frequencies to investigate what area of the vocal tract may be responsible for this change. Estimates using only F1—reflecting changes in the pharyngeal cavity—showed an average baseline VTL_{F1} of 13.99 cm, and an increase of 0.20 cm (SD: 0.59) per log(day) ($p < 0.01$) for an average VTL_{F1} of 15.20 cm after one year. Estimates using only F2—reflecting changes in the oral cavity—showed an average baseline VTL_{F2} of 13.69 cm, and an increase of 0.14 cm (SD: 0.20) per log(day) ($p < 0.001$). Estimates using only F3—which has no direct interpretation with respect to vocal-tract dimensions—showed an average baseline of 14.76 cm, increasing by 0.03 cm (SD: 0.16) per log(day). This is the only VTL estimate that was calculated that did not demonstrate a statistically significant change, though it was approaching statistical significance at a $p = 0.053$. See Table 6 for more details of the regression analysis. The results for MF0 from all groups can be found in Table 7. Individual regression results for all 60 participants in the trans group can be found in Table 8.

	Vocal Tract Length			
	From F1	**From F2**	**From F3**	**AVG**
Mean β (SD)	0.21 (0.59) cm/log(day)	0.14 (0.20) cm/log(day)	0.03(0.16) cm/log(day)	0.13 (0.25) cm/log(day)
Standard error	0.59	0.03	0.02	0.03
t	2.76	5.45	1.64	3.99
p	0.004 **	< 0.001 ***	0.053 *	< 0.001 ***
Mean at day 0	13.97 cm	13.69 cm	14.76 cm	14.14 cm
Mean at day 365	15.21 cm	14.53 cm	14.96 cm	14.90 cm
Annual change	+1.24 cm	+0.84 cm	+0.20 cm	+0.76 cm
% Annual change	8.88 %	6.14 %	1.36 %	5.37 %

Table 6 Regression analysis results for vocal tract length. F1: first formant. F2: second formant. F3: third formant. AVG: average estimate from F1, F2, and F3. * indicates approaching statistical significance of $p < 0.05$. ** indicates statistical significance at $p < 0.01$. *** indicates statistical significance at $p < 0.001$.

		Vocal Tract Length (cm)							
		From F1		From F2		From F3		AVG	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Trans group	Transmasc people (n = 60)	15.21	3.00	14.53	1.06	14.96	0.66	14.90	1.04
Control groups	Cis men (n = 11)	14.45	1.17	14.87	0.47	15.74	0.44	15.02	0.61
	Cis women (n = 17)	14.19	1.22	13.47	0.44	14.52	0.42	14.06	0.58
TIMIT database	Cis men (n = 112)	13.86	0.70	15.75	0.55	16.92	0.67	15.51	0.46
	Cis women (n = 56)	12.20	0.73	13.55	0.48	14.68	0.53	13.48	0.48

Table 7 Results for vocal tract length. Transmasc values reflect results after one year on T. VTL: vocal tract length. SD: standard deviation. F1: first formant. F2: second formant. F3: third formant. AVG: average estimate from F1, F2, and F3.

Participant number	Projected VTL after one year (cm)			
	From F1	From F2	From F3	AVG
1	11.54	14.72	15.20	13.82
2	15.74	17.34	15.58	16.22
3	8.36	12.05	12.94	11.12
4	14.63	15.11	14.83	14.86
5	19.65	16.09	15.84	17.19
6	15.51	13.92	14.76	14.73
7	16.47	16.00	15.58	16.02
8	17.00	14.46	14.16	15.20
9	14.90	16.73	14.90	15.51
10	13.75	12.88	14.71	13.78
11	16.64	13.36	15.62	15.21
12	15.44	15.10	14.40	14.98
13	18.83	14.03	14.75	15.87
14	18.70	14.50	15.31	16.17
15	14.09	14.81	14.79	14.56
16	16.69	13.99	14.45	15.05
17	15.63	15.27	15.26	15.39
18	18.01	13.67	14.90	15.53
19	13.78	15.16	15.63	14.86
20	11.31	15.95	13.52	13.59

Participant number	Projected VTL after one year (cm)			
	From F1	From F2	From F3	AVG
21	13.14	13.81	15.20	14.05
22	13.88	13.36	15.42	14.22
23	16.49	13.75	15.10	15.11
24	15.66	14.06	15.30	15.01
25	15.80	14.21	14.03	14.68
26	14.89	13.80	14.79	14.50
27	18.46	13.31	15.80	15.85
28	13.40	15.02	14.49	14.31
29	15.09	14.98	15.07	15.05
30	14.89	14.29	15.29	14.82
31	18.00	14.37	15.83	16.07
32	16.52	14.54	15.08	15.38
33	15.09	15.48	15.39	15.32
34	14.48	15.71	15.20	15.13
35	12.00	13.24	13.36	12.87
36	16.23	15.25	15.21	15.56
37	15.48	13.17	15.13	14.60
38	15.39	13.81	15.84	15.01
39	11.12	14.32	14.41	13.28
40	12.80	14.91	15.08	14.27
41	15.43	13.77	14.87	14.69
42	13.54	13.00	13.83	13.46
43	21.13	14.12	15.73	16.99
44	15.21	15.04	15.46	15.24
45	13.59	14.16	14.89	14.21
46	15.46	13.70	14.15	14.44
47	15.53	15.01	15.75	15.43
48	14.52	13.73	14.82	14.36
49	20.98	15.06	16.19	17.41
50	15.80	16.63	15.44	15.96
51	14.28	13.57	13.69	13.85
52	14.28	14.13	14.81	14.41
53	13.67	16.64	15.72	15.34
54	15.03	16.28	14.70	15.34
55	17.61	14.16	14.95	15.58
56	11.49	13.94	14.42	13.29
57	14.84	15.39	14.92	15.05
58	15.40	13.73	14.63	14.59
59	15.74	15.17	15.50	15.47
60	13.54	14.21	14.79	14.18

Table 8 Projected vocal tract length after one year. VTL: vocal tract length. F1: first formant. F2: second formant. F3: third formant. AVG: average estimate from F1, F2, and F3

Due to the large heterogeneity in recording schedules among the YouTube users, data was binned according to month for visualization. Binning by month was done according to the conversion found in Table 2. Results for MF0 and VTL were binned by month, and group averages for each month were plotted to visualize change over time for the YouTube group (Figure 7 and Figure 8). Mean and median values for F0 and VTL, binned by month, can be viewed in Table 9. A normal distribution curve of VTL from the trans group and the cis men and women from the TIMIT data is shown in Figure 9.

.

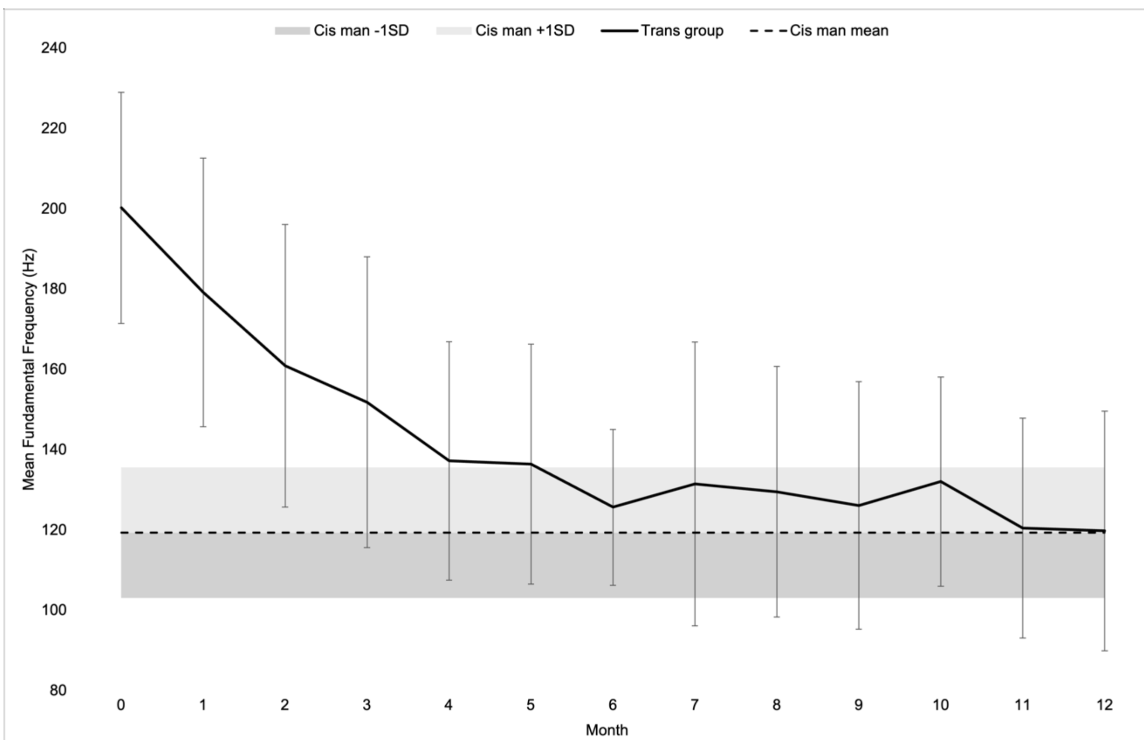

Figure 7 Mean fundamental frequency vs time (binned by month)

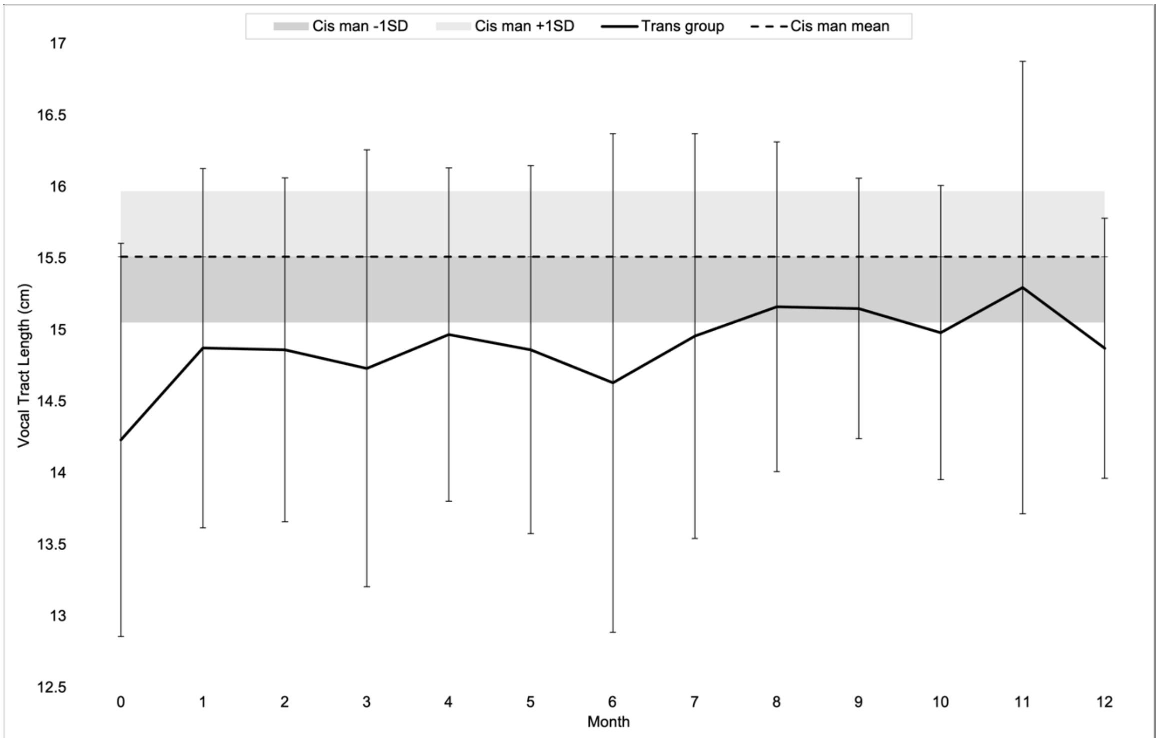

Figure 8 Vocal tract length vs time (binned by month)

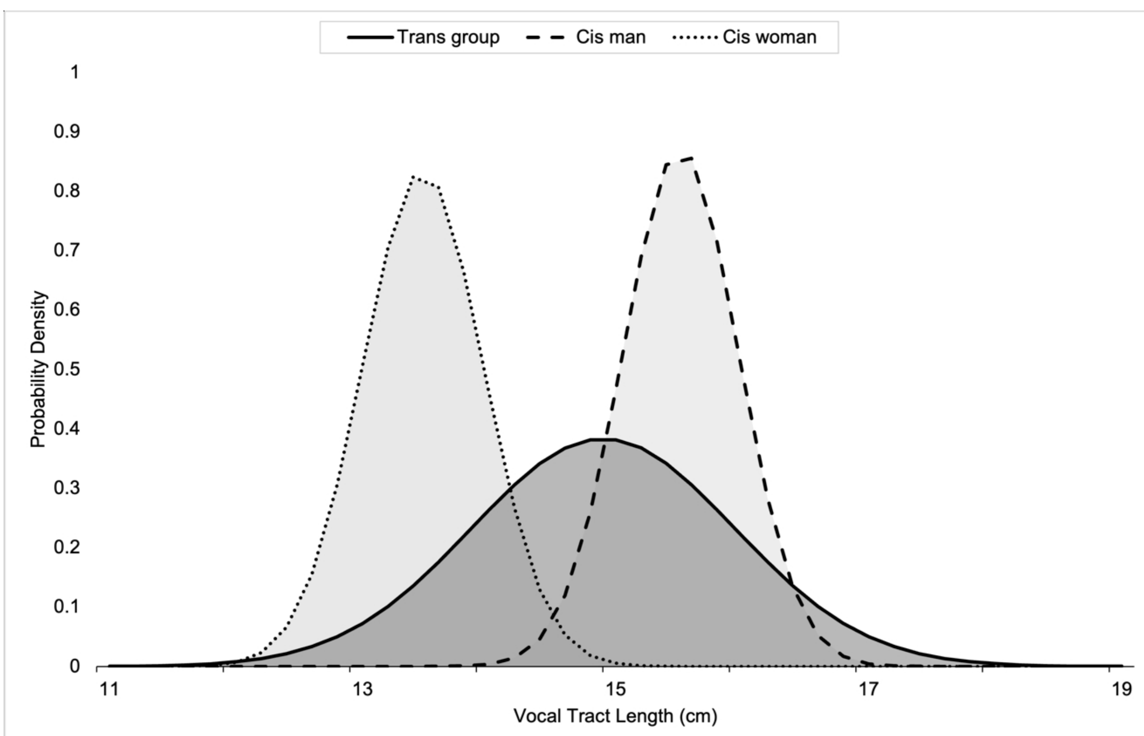

Figure 9 Distribution of vocal tract length

Month	F0 (Hz)		VTL (cm)							
			From F1		From F2		From F3		AVG	
	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med
0	200.68	195.53	14.09	14.11	13.76	13.69	14.85	14.74	14.23	14.20
1	179.19	177.09	14.97	14.44	14.76	14.44	14.89	14.83	14.87	14.70
2	160.91	153.89	15.30	15.17	14.37	14.33	14.92	14.91	14.86	14.94
3	151.85	144.47	15.07	15.36	14.22	14.12	14.91	15.00	14.73	14.86
4	137.23	131.02	15.09	14.67	14.84	14.49	14.98	15.08	14.97	15.06
5	136.43	125.40	14.97	14.39	14.63	14.45	14.99	15.05	14.86	14.54
6	125.70	119.29	14.89	15.23	14.19	14.08	14.81	14.90	14.63	14.70
7	131.49	120.69	14.97	14.94	15.01	14.92	14.90	14.89	14.96	14.95
8	129.55	120.98	15.79	15.60	14.61	14.78	15.08	15.13	15.16	15.20
9	126.14	117.63	15.30	15.12	15.05	14.90	15.10	15.13	15.15	15.11
10	132.09	122.77	15.19	14.79	14.83	14.74	14.93	15.16	14.98	14.76
11	120.49	113.80	15.59	15.57	15.41	15.06	14.89	14.95	15.30	15.40
12	119.79	109.72	15.12	14.93	14.44	14.51	15.05	15.12	14.87	14.76

Table 9 Binned values for mean fundamental frequency and vocal tract length. F0: fundamental frequency. VTL: vocal tract length. F1: first formant. F2: second formant. F3: third formant. AVG: average estimate from F1, F2, and F3.

Chapter 4: Discussion

4.1 How Do Speech Acoustics Change in One Year?

4.1.1 Fundamental Frequency

As anticipated, the results from the YouTube control group demonstrate that fundamental frequency is significantly reduced after one year on T. Confirming what has been noted previously in the literature, this study supports the notion that T is effective at reducing a speaker's MF0 to fall within one standard deviation of that of the average cis man.

4.1.2 Vocal Tract Length

Echoing the work done previously by Cler et al. (2020) and Hodges-Simeon et al. (2021), this study found that there was a significant change in VTL after one year on T. The resulting VTL in the trans group fell within two standard deviations of the average cis man's VTL. Additionally, 80% (48 out of 60) trans speakers had a resulting VTL that fell within the range of the cis men (14.03cm to 16.77 cm), while at baseline, only 53% (32 out of 60) of the trans speakers had a VTL that fell within that range.

4.1.2.1 Where is the Vocal Tract Growing?

After discovering the increase in VTL with an average estimate from F1, F2, and F3, further calculations were performed to separate the effect of F1 and F2 on the VTL increase. This allowed us to isolate which area of the vocal tract is experiencing a higher rate of growth. If a regression on VTL estimates from F1 showed a greater average regression coefficient, then we can posit that the pharyngeal cavity is growing more rapidly than the oral cavity. If F2 showed a greater rate, then the oral cavity would be growing faster. Our results show that the pharyngeal cavity was growing at a faster rate than the oral cavity.

The mechanism of action behind the growth of the pharyngeal cavity remains a question to be answered. Two possible hypotheses have been considered. First, the pharyngeal cavity could, itself, be undergoing structural changes. This could be due to physical growth that is causing the neck that houses the pharynx to become longer, perhaps from an increase in overall body height. However, one would not expect any increase in body height once puberty is complete, even with T. Additionally, there is a lack of studies investigating height increases resulting from T, especially in clients starting T as an adult. Assuming the speakers in the trans group were all post-pubertal at the time their videos were recorded, this hypothesis lacks support.

The second hypothesis, which is the one that the author feels is more likely, is that the pharyngeal cavity has lengthened because the speakers are lowering their larynx during speech. As noted in the introduction section, speakers can manipulate various structures involved in speech to modify their voice acoustics. This includes adjusting the position of the larynx to sit higher or lower in the pharynx. This practice is used in the clinical setting when trans clients are involved in gender affirming voice training with a speech-language pathologist (SLP). It is possible that the trans speakers are naturally speaking with a laryngeal position that is lower than their pre-T baseline, either consciously or subconsciously, based on auditory feedback they receive as they speak. Simply put, they may find they sound better that way. Studies using imaging or endoscopy are required to verify the root of this pharyngeal change and test each of these hypotheses.

4.2 Clinical Relevance

Developing a possible timeline for voice changes in trans clients taking T is a powerful clinical tool for multiple professions. Primary-care physicians and endocrinologists who work with trans clients taking T as a form of HRT can use this information to provide effective counselling and a more accurate timeline for changes their clients might expect to see during their voice transitions.

In the field of speech-language pathology, clinicians can use this information to prioritize treatment goals. For instance, knowing that MF0 is likely to change to fall within a client's ideal range but VTL may not make as large of a change could mean that clinicians and clients can choose to focus on resonance goals. If the client is satisfied with their predicted VTL, therapy efforts can be focused on additional gender markers in the voice, such as articulatory patterns and prosodic features.

Another valuable use of this information is its potential for use in voice synthesis. SLPs can apply these mean values to the voices of their clients to synthesize a possible future voice. This can be useful for counselling as well as in therapy. For instance, a client in therapy can use their predicted future voice as a target to hit when performing pitch and resonance modifying activities. This may be a more effective tool than using a piano for target pitch, which is commonly used in gender affirming voice training today.

4.3 Future Research

Although this study was able to recruit the largest number of trans speakers to date, there is still a need for more research with larger pools of speakers to determine norms. These studies should, ideally, include details that were missing in this study, such as height, age, and details about dose.

Future research should include norm/control groups speaking the same standardized phrases as the test group. Ideally, future research with prospectively recruited participants would include sustained vowels from both test and control groups to capture formant frequencies more accurately. Additionally, research investigating the differences in response to treatment between those who started testosterone in adolescence when puberty has not been completed and those who started after completing puberty. Results of this area of research could support the value of continuing to offer gender affirming hormones to youth.

After confirming what two previous studies had uncovered about increases in VTL in this population, this study provides a rationale for further investigation into this effect of T. In particular, studies using imaging or endoscopic measures may shine some light on the source of changes in the vocal tract.

4.5 Limitations

Due to this study's retrospective design that uses data found online, there are a number of notable limitations to mention. Firstly, there is bias in the selection process resulting from the nature of social media websites' search engines which use algorithms to cater search results to each individual user. Additionally, YouTube, similar to other social media sites, ranks search results based on metrics that are potentially damaging to performing an objective or random selection, such as relevance, video view count, and creator subscriber count. To avoid the algorithmic bias of search results, specific search strategies, noted in Sui et al. (2022), were employed. This included using the incognito mode of Google Chrome which removes cookies, thus eliminating the website's ability to personalize search results, and conducting multiple searches that order the results in different ways (i.e., by relevance, by view count, and by subscriber count) to reduce the impact of ordering on the selection of videos. What could not be avoided, however, is self-selection bias which may lead people to only post their voice results if there were impressive changes and the poster is highly satisfied with their voice.

A second limitation is related to the lack of personal information known about the YouTube participants. Details such as age, dose, route of administration, and access to hormone blockers as an adolescent are all things that could impact someone's response to T. Other details, such as height, race, pre-existing conditions, pre-treatment testosterone serum levels, and participation in voice therapy, could also influence the voice and explain some variance between test subjects. Additionally, it is entirely possible that some videos contained audio that was digitally modified or produced by artificial intelligence. Unfortunately, due to the recruitment method of using publicly available videos, personal details are not accessible and could not be included in the study, and the authenticity of the data cannot be verified. Further research with strategies that allow access to personal details on participants is needed, as well as prospective studies that could control for more factors than was feasible in this study.

Another limitation is the variance in the recordings. Each participant recorded themselves for different periods of treatment time and at different intervals. Additionally, some individuals recorded themselves with better quality equipment or in better recording environments with lower levels of ambient noise that may have impacted acoustic analysis. This was mitigated as best as possible by screening out videos that included perceivable levels of background noise. However, this may have biased the results as it eliminated many recordings. In a similar vein, not all YouTube users recited the standardized phrase during their recordings. All speech spoken during a specific treatment period was included in the analysis to avoid biased selection of the sample. Some people spoke more than others, and perhaps their voice then reflected a speaking F0, while others recited short, standardized sentences, that may be considered a reading F0. Literature has previously made a distinction between these two varieties of F0, so this might have impacted results.

The goal with this study was to include control groups of cis men and women speaking the same phrase as the study group for comparison after one year. However, failure to recruit enough control participants lead to the abandonment of this plan. In replacement, the results from the test group were compared to results from an existing database of continuous speech. While method of extracting data from all groups were identical, the difference in terms of the content spoken by the trans group and the TIMIT group means that formant frequencies may have been impacted by the difference in vowels present. This is not ideal for comparison.

Finally, the formant ceiling cut-off of 5500 Hz used for the trans group may have led to the formant frequency results being positively skewed. This limitation is unavoidable without manually selecting a ceiling for every speaker in all groups. Selecting the higher ceiling value, rather than 5000 Hz suggested for cis men, was the more conservative choice, and was necessary to automate the process of formant frequency extraction, considering the quantity of recordings to be analyzed. However, this means that formant values may not reflect the true VTL of the trans group. More accurate values may indeed be lower than reported, reflecting even greater VTL increases.

References

Amine, R., Gonzalez, R. G., Hagemeister, P., Valsorda, F., Ferrándiz, J. M., Hsuan, Y. C., Jonaitis, N., pulpe, & M, S. (2021). Youtube-dl [Computer software]. https://github.com/ytdl-org/youtube-dl

Boersma, P., & Weenick, D. (2023). Praat: Doing phonetics by computer (6.3.20) [Computer software]. https://fon.hum.uva.nl/praat/

Borck, C. R., & Moore, L. J. (2019). This is my voice on T: synthetic testosterone, DIY surveillance, and transnormative masculinity. Surveillance & Society, 17(5), 631–640. https://doi.org/10.24908/ss.v17i5.12931

Brown, W. S., Morris, R. J., Hicks, D. M., & Howell, E. (1993). Phonational profiles of female professional singers and nonsingers. Journal of Voice, 7(3), 219–226. https://doi.org/10.1016/S0892-1997(05)80330-3

Cler, G. J., McKenna, V. S., Dahl, K. L., & Stepp, C. E. (2020). Longitudinal case study of transgender voice changes under testosterone hormone therapy. Journal of Voice, 34(5), 748–762. https://doi.org/10.1016/j.jvoice.2019.03.006

Damrose, E. J. (2009). Quantifying the impact of androgen therapy on the female larynx. Auris Nasus Larynx, 36(1), 110–112. https://doi.org/10.1016/j.anl.2008.03.002

Deuster, D., Di Vincenzo, K., Szukaj, M., Am Zehnhoff-Dinnesen, A., & 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. https://doi.org/10.1007/s00405-016-4043-0

Deuster, D., Matulat, P., Knief, A., Zitzmann, M., Rosslau, K., Szukaj, M., Am Zehnhoff-Dinnesen, A., & Schmidt, C.-M. (2016). Voice deepening under testosterone treatment in female-to-male gender dysphoric individuals. European Archives of Oto-Rhino-Laryngology, 273(4), 959–965. https://doi.org/10.1007/s00405-015-3846-8

Fairbanks, G. (1960). The Rainbow Passage. In Voice and articulation drillbook (2nd ed., pp. 124–139). Harper & Row.

Ferguson, S. (2021, January 13). What Does it Mean to be Transmasculine? Healthline. https://www.healthline.com/health/transmasculine

Fitch, J. L., & Holbrook, A. (1970). Modal vocal fundamental frequency of young adults. Archives of Otolaryngology - Head and Neck Surgery, 92(4), 379–382. https://doi.org/10.1001/archotol.1970.04310040067012

Fulop, S. A. (2011). Introduction. In S. A. Fulop, Speech Spectrum Analysis (pp. 1–4). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17478-0_1

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Pallett, D. S., Dahlgren, N. L., Zue, V., & Fiscus, J. G. (1993). TIMIT Acoustic-Phonetic Continuous Speech Corpus. [dataset]. https://doi.org/10.35111/17GK-BN40

Gelfer, M. P., & Bennett, Q. E. (2013). Speaking fundamental frequency and vowel formant frequencies: Effects on perception of gender. Journal of Voice, 27(5), 556–566. https://doi.org/10.1016/j.jvoice.2012.11.008

Goldstein, U. G. (1980). An articulatory model for the vocal tracts of growing children [Doctoral dissertation]. Massachusetts Institute of Technology.

Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., & Van Lieshout, P. (2013). Normative voice data for younger and older adults. Journal of Voice, 27(5), 545–555. https://doi.org/10.1016/j.jvoice.2013.03.002

Groll, M. D., Dahl, K. L., C, ádiz M. D., Welch, B., Tracy, L. F., & Stepp, C. E. (2022). Resynthesis of transmasculine voices to assess gender perception as a function of testosterone therapy. Journal of Speech, Language, and Hearing Research, 65(7), 2474–2489. https://doi.org/10.1044/2022_JSLHR-21-00482

Hancock, A. B., Childs, K. D., & Irwig, M. S. (2017). Trans male voice in the first year of testosterone therapy: Make no assumptions. Journal of Speech, Language, and Hearing Research, 60(9), 2472–2482. https://doi.org/10.1044/2017_JSLHR-S-16-0320

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. Attention, Perception, & Psychophysics, 71(5), 1150–1166. https://doi.org/10.3758/APP.71.5.1150

Hodges-Simeon, C. R., Grail, G. P. O., Albert, G., Groll, M. D., Stepp, C. E., Carré, J. M., & Arnocky, S. A. (2021). Testosterone therapy masculinizes speech and gender presentation in transgender men. Scientific Reports, 11(1), 3494. https://doi.org/10.1038/s41598-021-82134-2

Howard, D. M., & Murphy, D. (2008). Voice Science, Acoustics and Recording. Plural Pub.

Irwig, M. S., Childs, K., & Hancock, A. B. (2017). Effects of testosterone on the transgender male voice. Andrology, 5(1), 107–112. https://doi.org/10.1111/andr.12278

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI): Development and validation. American Journal of Speech-Language Pathology, 6(3), 66–70. https://doi.org/10.1044/1058-0360.0603.66

Lennes, M. (2003). SpeCT - The Speech Corpus Toolkit for Praat. https://lennes.github.io/spect/

Leung, Y., Oates, J., & Chan, S. P. (2018). Voice, articulation, and prosody contribute to listener perceptions of speaker gender: A systematic review and meta-analysis. Journal of Speech, Language, and Hearing Research, 61(2), 266–297. https://doi.org/10.1044/2017_JSLHR-S-17-0067

Markova, D., Richer, L., Pangelinan, M., Schwartz, D. H., Leonard, G., Perron, M., Pike, G. B., Veillette, S., Chakravarty, M. M., Pausova, Z., & Paus, T. (2016). Age- and sex-related variations in vocal-tract morphology and voice acoustics during adolescence. Hormones and Behavior, 81, 84–96. https://doi.org/10.1016/j.yhbeh.2016.03.001

Merritt, B., & Bent, T. (2024). Revisiting the acoustics of speaker gender perception: A gender expansive perspective. J. Acoust. Soc. Am.

Morris, R. J., Brown, W. S., Hicks, D. M., & Howell, E. (1995). Phonational profiles of male trained singers and nonsingers. Journal of Voice, 9(2), 142–148. https://doi.org/10.1016/S0892-1997(05)80247-4

Nygren, U., Nordenskjöld, A., Arver, S., & Södersten, M. (2016). Effects on voice fundamental frequency and satisfaction with voice in trans men during testosterone treatment—A longitudinal study. Journal of Voice, 30(6), 766.e23-766.e34. https://doi.org/10.1016/j.jvoice.2015.10.016

Planned Parenthood. (n.d.). Sex and Gender Identity. Planned Parenthood. https://www.plannedparenthood.org/learn/gender-identity/sex-gender-identity

Raun, T. (2015). Archiving the wonders of testosterone via YouTube. TSQ: Transgender Studies Quarterly, 2(4), 701–709. https://doi.org/10.1215/23289252-3151646

Sherbourne Health. (2020). https://www.rainbowhealthontario.ca/TransHealthGuide/intro-terms.html

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. Journal of Speech, Language, and Hearing Research, 57(1), 285–296. https://doi.org/10.1044/1092-4388(2013/12-0314)

Sui, W., Sui, A., & Rhodes, R. E. (2022). What to watch: Practical considerations and strategies for using YouTube for research. Digital Health, 8, 205520762211237. https://doi.org/10.1177/20552076221123707

Sun, S. D. (2019, June 13). Stop Using Phony Science to Justify Transphobia. Scientific American. https://www.scientificamerican.com/blog/voices/stop-using-phony-science-to-justify-transphobia/

Top Websites Ranking. (2025). Similarweb. https://www.similarweb.com/top-websites/

Van Borsel, J., De Cuypere, G., Rubens, R., & Destaerke, B. (2000). Voice problems in female□to□male transsexuals. International Journal of Language & Communication Disorders, 35(3), 427–442. https://doi.org/10.1080/136828200410672

Zamponi, V., Mazzilli, R., Mazzilli, F., & Fantini, M. (2021). Effect of sex hormones on human voice physiology: From childhood to senescence. Hormones, 20(4), 691–696. https://doi.org/10.1007/s42000-021-00298-y

Ziegler, A., Henke, T., Wiedrick, J., & Helou, L. B. (2018). Effectiveness of testosterone therapy for masculinizing voice in transgender patients: A meta-analytic review. International Journal of Transgenderism, 19(1), 25–45. https://doi.org/10.1080/15532739.2017.1411857

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., Thrush, C. R., & Glaze, L. E. (2011). Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). American Journal of Speech-Language Pathology, 20(1), 14–22. https://doi.org/10.1044/1058-0360(2010/09-0105)

Appendix A: Screening Questionnaire for Control Group

Who Can Take Part in the Research Study

You may participate as a member of a control group in this study if you are a cisgender man or cisgender woman who is between the ages of 18 and 35 who has no history of voice disorders or hormone disorders. Cisgender means that your gender matches your sex registered at birth. These details have been described in the email previously sent to you. They are repeated below for your review.

If **any** of the following statements apply to you, you are **not eligible** to participate in the study:

1. I am younger than 18 or older than 35.
2. I am transgender.
3. I am intersex or have a condition that falls under the intersex umbrella (e.g., congenital adrenal hyperplasia, androgen insensitivity syndrome, gonadal dysgenesis, Turner syndrome, Klinefelter syndrome).
4. I have a sex hormone disorder/abnormal sex hormone levels (e.g., polycystic ovarian syndrome, low testosterone, hypopituitarism, hypogonadism, androgen insufficiency).
5. I have a hormone disorder that may impact the voice (e.g., Addison disease, Cushing syndrome, Cushing disease, growth hormone deficiency, hypothyroidism, acromegaly)
6. My puberty was delayed (i.e., as a cisgender man, there were no signs of testicular development by 14 years of age; as a cisgender woman, there were no signs of breast development by 13 years of age OR there was breast development but no menstrual period by 15 years of age).
7. My puberty was premature (i.e., as a cisgender man, there were signs of puberty before 9 years of age; as a cisgender woman, there were signs of puberty before 8 years of age).
8. I have taken supplemental testosterone as prescribed by a doctor.
9. I have taken anabolic steroids for medical or personal use (e.g., nandrolone (Deca-Durabolin), oxandrolone (Oxandrin)).

10. I take or have taken one of the following medications: danazol (Danocrine), flutamide (Eulexin), spironolactone (Aldactone), cyproterone (Diane or Diane-35).
11. I have and/or used to have a voice disorder (e.g., dysphonia, vocal nodes, vocal cord paralysis or paresis, puberphonia).

If **two or more** of the below statements apply to you, you are **not eligible** to participate.

Physical measures from the Voice Handicap Index (Jacobson et al., 1997).

Each measure rated: yes/no

1. I run out of air when I talk.
2. The sound of my voice varies throughout the day.
3. People ask, "What's wrong with your voice?"
4. My voice sounds creaky and dry.
5. I feel as though I have to strain to produce voice.
6. The clarity of my voice is unpredictable.
7. I try to change my voice to sound different.
8. I use a great deal of effort to speak.
9. My voice is worse in the evening.
10. My voice "gives out" on me in the middle of speaking.

Appendix B: The Rainbow Passage segment

"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." From Fairbanks (1960).

Appendix C: Praat Scripts

Formants

```
# This script goes through sound and TextGrid files in a
directory,
# opens each pair of Sound and TextGrid, calculates the
formant values
# at the midpoint of each labeled interval, and saves
results to a text file.
# To make some other or additional analyses, you can modify
the script
# yourself... it should be reasonably well commented! ;)
#
# This script is distributed under the GNU General Public
License.
# 4.7.2003 Mietta Lennes

form Analyze formant values from labeled segments in files
     comment Which tier do you want to analyze?
     sentence Tier Mary
     comment Formant analysis parameters
     positive Time_step 0.01
     integer Maximum_number_of_formants 5
     positive Maximum_formant_(Hz) 5500
     positive Window_length_(s) 0.025
     real Preemphasis_from_(Hz) 50
    comment Filename extensions for the files to be
analyzed:
     sentence Sound_file_extension .wav
     sentence TextGrid_file_extension .TextGrid
endform

writeInfoLine: "- Select the directory with the sound files
to be analyzed"
sound_directory$ = chooseDirectory$: "Select the directory
with the sound files to be analyzed"
appendInfoLine: "- Select the directory with the TextGrid
files to be analyzed"
textGrid_directory$ = chooseDirectory$: "Select the
directory with the TextGrid files to be analyzed"

resultfile$ = chooseWriteFile$: "Save the analysis results
to file:", "formantresults.txt"
```

```
# Here, you make a listing of all the sound files in a
directory.

fileList = Create Strings as file list: "Files in
directory", sound_directory$ +"/*" + sound_file_extension$

numberOfFiles = Get number of strings
appendInfoLine: "'newline$''numberOfFiles' sound files
found in 'sound_directory$'"

# Check if the result file exists:
if fileReadable (resultfile$)
     # pause The result file 'resultfile$' already exists!
Do you want to overwrite it?
     deleteFile: resultfile$
endif

# Write a row with column titles to the result file:
# (remember to edit this if you add or change the
analyses!)

titleline$ = "Filename" + tab$ + "Segment label" + tab$ +
"F1 (Hz)" + tab$ + tab$ + "F2 (Hz)" + tab$ + tab$ + "F3
(Hz)"
appendFileLine: resultfile$, titleline$

# Go through all the sound files, one by one:

for ifile to numberOfFiles
          filename$ = Get string: ifile
          # A sound file is opened from the listing:
          sound = Read from file: sound_directory$ + "/" +
filename$
     # Starting from here, you can add everything that
should be
          # repeated for every sound file that was opened:
          soundname$ = selected$ ("Sound", 1)
          To Formant (burg)... time_step
maximum_number_of_formants maximum_formant window_length
preemphasis_from
          # Open a TextGrid by the same name:
          gridfile$ = textGrid_directory$ + "/" +
soundname$ + textGrid_file_extension$
          if fileReadable (gridfile$)
               Read from file... 'gridfile$'
```

```
		# Find the tier number that has the label
given in the form:
		call GetTier 'tier$' tier
		numberOfIntervals = Get number of
intervals... tier
		# Pass through all intervals in the selected
tier:
		for interval to numberOfIntervals
			label$ = Get label of interval... tier
interval

			if label$ <> ""
				# if the interval has an unempty
label, get its start and end:
				start = Get starting point... tier
interval
				end = Get end point... tier
interval
				midpoint = (start + end) / 2
				# get the formant values at that
interval
				select Formant 'soundname$'
				f1 = Get quantile... 1 start end
Hertz 0.5
				f2 = Get quantile... 2 start end
Hertz 0.5
				f3 = Get quantile... 3 start end
Hertz 0.5
				# Save result to text file:
				appendFileLine: resultfile$,
soundname$ + tab$ + label$ + tab$ + tab$ + fixed$ (f1, 3) +
tab$ + tab$ + fixed$ (f2, 3) + tab$ + tab$ + fixed$ (f3, 3)
				select TextGrid 'soundname$'
			endif
		endfor
		# Remove the TextGrid object from the object list
		select TextGrid 'soundname$'
		Remove
	endif
	# Remove the temporary objects from the object list
	select Sound 'soundname$'
	plus Formant 'soundname$'
	Remove
	selectObject: fileList
	# and go on with the next sound file!
endfor
```

```
Remove

appendInfoLine: "'newline$'Finished!"
appendInfoLine: "'newline$'The analysis results were saved
to this file: 'newline$'  'resultfile$'"

#-------------
# This procedure finds the number of a tier that has a
given label.

procedure GetTier name$ variable$
        numberOfTiers = Get number of tiers
        itier = 1
        repeat
                tier$ = Get tier name... itier
                itier = itier + 1
        until tier$ = name$ or itier > numberOfTiers
        if tier$ <> name$
                'variable$' = 0
        else
                'variable$' = itier - 1
        endif

     if 'variable$' = 0
          exit The tier called 'name$' is missing from the
file 'soundname$'!
     endif

endproc
```

Fundamental Frequency

```
# This script goes through sound and TextGrid files in a
directory,
# opens each pair of Sound and TextGrid, calculates the
pitch maximum
# of each labeled interval, and saves results to a text
file.
# To make some other or additional analyses, you can modify
the script
# yourself... it should be reasonably well commented! ;)
#
```

```
# This script is distributed under the GNU General Public
License.
# Copyright 4.7.2003 Mietta Lennes http://orcid.org/0000-
0003-4735-3017
# Praat new syntax version 2016-08-14 stefan.werner@uef.fi
# A few more slight improvements 2016-11-24 Mietta Lennes
http://orcid.org/0000-0003-4735-3017

form Analyze mean pitch from labeled segments in files
     comment Which tier do you want to analyze?
     sentence Tier Mary
     comment Pitch analysis parameters
     positive Time_step 0.01
     positive Minimum_pitch_(Hz) 75
     positive Maximum_pitch_(Hz) 400
     comment Filename extensions for the files to be
analyzed:
     sentence Sound_file_extension .wav
     sentence TextGrid_file_extension .TextGrid
endform

writeInfoLine: "- Select the directory with the sound files
to be analyzed"
sound_directory$ = chooseDirectory$: "Select the directory
with the sound files to be analyzed"
appendInfoLine: "- Select the directory with the TextGrid
files to be analyzed"
textGrid_directory$ = chooseDirectory$: "Select the
directory with the TextGrid files to be analyzed"

resultfile$ = chooseWriteFile$: "Save the analysis results
to file:", "pitchresults.txt"

# Here, you make a listing of all the sound files in a
directory.

fileList = Create Strings as file list: "Files in
directory", sound_directory$ +"/*" + sound_file_extension$

numberOfFiles = Get number of strings
appendInfoLine: "'newline$''numberOfFiles' sound files
found in 'sound_directory$'"

# Check if the result file exists:
if fileReadable (resultfile$)
```

```
  # uncomment next line if you want to stop the script
before it overwrites an old result file
     # pauseScript: "The result file ", resultfile$, "
already exists! Do you want to overwrite it?"
     deleteFile: resultfile$
endif

# Write a row with column titles to the result file:
# (remember to edit this if you add or change the
analyses!)

titleline$ = "Filename" + tab$ + "Segment label" +  tab$ +
     "Mean Pitch (Hz)"
appendFileLine: resultfile$, titleline$

# Go through all the sound files, one by one:

for ifile to numberOfFiles
     filename$ = Get string: ifile
     # A sound file is opened from the listing:
     sound = Read from file: sound_directory$ + "/" +
filename$
  # Starting from here, you can add everything that should
be
     # repeated for every sound file that was opened:
     soundname$ = selected$ ("Sound")
     pitch = To Pitch: 0.00, 75.00, 400.00
     # Open a TextGrid by the same name:
     gridfile$ = textGrid_directory$ + "/" + soundname$ +
textGrid_file_extension$
     if fileReadable (gridfile$)
          tg = Read from file: gridfile$
          # Find the tier number that has the label given
in the form:
          @getTier: tier$
          numberOfIntervals = Get number of intervals:
getTier.number
          # Pass through all intervals in the selected
tier:
          for interval to numberOfIntervals
               label$ = Get label of interval:
getTier.number, interval
               if label$ <> ""
                    # if the interval has an unempty label,
get its start and end:
```

```
                    start = Get starting point:
getTier.number, interval
                    end = Get end point: getTier.number,
interval
                    # get f0 at that interval
                    selectObject: pitch
                    f0 = Get mean: start, end, "Hertz"
                    # Save result to text file:
                    resultline$ = soundname$ + tab$ +
label$ + tab$ + fixed$ (f0, 3)
                    appendFileLine: resultfile$,
resultline$
                    selectObject: tg
               endif
          endfor
          # Remove the TextGrid object from the object list
          selectObject: tg
          Remove
     endif
     # Remove the temporary objects from the object list
     selectObject: sound
     plusObject: pitch
     Remove
     selectObject: fileList
     # and go on with the next sound file!
endfor

Remove

appendInfoLine: "'newline$'Finished!"
appendInfoLine: "'newline$'The analysis results were saved
to this file: 'newline$'  'resultfile$'"


#-------------
# This procedure finds the number of a tier that has a
given name.
# (There is no real need for a proc here since the code is
only run once
# - but you may find the snippet useful in your own
scripts.)

procedure getTier: .tiername$
        numberOfTiers = Get number of tiers
        itier = 1
        .number = 0
```

```
		repeat
			tryTier$ = Get tier name: itier
			itier = itier + 1
		until tryTier$ = tier$ or itier > numberOfTiers
		if tryTier$ <> tier$
			.number = 0
		else
			.number = itier - 1
		endif

	if .number = 0
		exitScript: "The tier called ",  tier$,  " is
missing from the file ",  soundname$, "!"
	endif

endproc
```