

Longitudinal Case Study of Transgender Voice Changes Under Testosterone Hormone Therapy

*,[†]Gabriel J. Cler, [†]Victoria S. McKenna, [†]Kimberly L. Dahl, and *,[†],[‡],^{||}Cara E. Stepp, *,[†] *|| Boston, Massachusetts*

Abstract: The purpose of this study was to comprehensively evaluate voice and speech changes in one healthy 30-year-old transgender male undergoing testosterone therapy for transition. Testing occurred at three timepoints before cross-sex hormone therapy and every 2 weeks thereafter for 1 year. Data collected included measures of acoustics, aerodynamics, and laryngeal structure and function via flexible laryngoscopy. Analysis included acoustic correlates of pitch, loudness, voice quality, and vocal tract length, as well as perceptual measures of voice quality and gender. Speaking fundamental frequency (f_0) lowered from 183 Hz to 134 Hz. Phonatory frequency range (ie, minimum and maximum singing range) shifted from a range of D#3–E6 to a range of A2–A5. Perceptual measures of voice quality indicated no negative changes. Naïve listeners reliably rated the participant's speech samples as male after 37 weeks on testosterone. Few studies document in detail the variety of voice changes that occur during cross-sex hormone therapy, focusing instead on f_0 alone. This study adds to the literature a comprehensive case study of speech and voice changes experienced by one transmasculine participant undergoing testosterone therapy.

Key Words: Transgender—Transmasculine—Endoscopy—Aerodynamics—Acoustic—HRT—Testosterone.

INTRODUCTION

Transgender voice has become highly relevant in speech-language pathology clinical practice, but research in this area is still lacking.¹ Of the transgender voice studies available, most focus on transfeminine voice, with fewer studies evaluating transmasculine voice.^{1,2} This focus has been explained by the fact that transmasculine people typically are not seen clinically for voice masculinization. The relative absence of transmasculine individuals from voice clinics is likely a multifactorial issue, which include psychosocial and socio-political factors.³ Within the speech-language pathology field, many clinicians report a lingering belief that masculinization is accomplished solely by the effects of testosterone on the voice.^{2,4} In contrast, estrogen therapy does not feminize the voice, so transfeminine clients are often seen in the voice clinic.⁴ The effects of testosterone therapy (ie, hormone replacement therapy or HRT) on all aspects of transmasculine voice are not yet characterized, and in fact testosterone therapy alone may not result in voice satisfaction.^{5–7} Transmasculine speakers experience a range of gender-related voice problems,⁸ which may or may not be resolved with hormone therapy and speech therapy.

There are limited studies investigating transmasculine voice,^{5–7,9–17} many of which only consider one or two domains of voice – typically focusing on fundamental frequency (f_0), which is perceived as pitch. Some studies suggest that speakers may experience changes and restrictions

in a variety of voice domains, including: pitch range/variability, vocal control/stability, glottal function, and voice quality.^{16,18} Voice quality in particular has been suggested to be reduced following testosterone therapy,^{9,12,16} but some evidence suggests that, while variable, it usually remains within normal limits.¹⁶ An even smaller set of studies have tracked *longitudinal* voice changes in trans men undergoing testosterone therapy (summarized in Table 1).^{5,7,12–14,17} These studies typically characterize speaking f_0 ^{5,7,12–14,17} during sustained vowels or reading, and occasionally assess phonatory frequency range (PFR)^{5,7,12,16} (that is, the entire range of singing f_0 that can be produced) and acoustic measures including jitter,^{7,16} shimmer,^{7,12,16} and noise-to-harmonic ratio.^{12,16} Some studies have also measured self-perception of voice.^{5,16}

The scarcity of research on transmasculine voice, and particularly the scarcity of evidence tracking speakers longitudinally, prevents endocrinologists and speech-language pathologists from providing transmasculine speakers with accurate prognoses of voice changes under testosterone (“T”). Understanding which voice features may be affected, over what period, and to what degree informs decisions regarding whether and when other approaches to voice masculinization, such as voice therapy, may be warranted. Further, given the diversity of voice-related goals among transmasculine speakers,⁸ a fuller understanding of expected changes is necessary to best serve this population with an evidence-based approach.

Here we present, following case report (CARE) guidelines,¹⁹ an in-depth case study of one transmasculine speaker's voice changes under testosterone. Analyses follow one speaker from before testosterone therapy (baseline) to 12 months on testosterone with twice monthly comprehensive data collection enabling videoendoscopic assessment, aerodynamic measures, acoustic measures, clinician perception of voice changes, and listener perception of speaker gender.

Accepted for publication March 14, 2019.

From the *Graduate Program for Neuroscience, Boston University, Boston, Massachusetts; †Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts; ‡Department of Biomedical Engineering, Boston University, Boston, Massachusetts; and the ||Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, Massachusetts.

Address correspondence and reprint requests to Gabriel J. Cler, Graduate Program for Neuroscience, Boston University, Boston, MA 02215. E-mail: gcler@bu.edu

Journal of Voice, Vol. ■■■, No. ■■■, pp. ■■■–■■■
0892-1997

© 2019 The Voice Foundation. Published by Elsevier Inc. All rights reserved.
<https://doi.org/10.1016/j.jvoice.2019.03.006>

METHODS

Participant, study design, and dosage

The participant was a 30-year-old transmasculine individual (assigned female at birth) who worked as a researcher. As such, he actively contributed to the design of the study (eg, suggesting measures of interest and negotiating the schedule of data collection), in an example of participatory action research.²⁰ He was a heavy voice user at work as well as an avocational but high-intensity (6 hours/week) choral singer. He had no history of speech, language, or voice issues and had not completed any voice therapy either prior to or during the course of the study. His gender identity was self-reported as male and he used *he/him/his* pronouns. He was receiving regular medical care from a gender-affirming medical clinic, and was in the process of medical testing before starting testosterone when he first participated in this study. He had begun socially transitioning 5 months before and had not yet begun the legal processes involved with changing his name and gender on governmental identification. He completed his name change around week 6 of this study and aligned other identification over the following months. He began pursuing gender-affirming surgery around week 20 of the study and had that surgery during week 32. Accordingly, he wore a chest binder for all recording sessions in the baselines through week 35, after which he was healed sufficiently to not require a binder or dressings.

He completed weekly intramuscular injections of testosterone cypionate for the purpose of masculinization. Dosage was set at 40 mg per injection (0.2 mL of 200 mg/mL solution) for weeks 1–42 and was increased to 60 mg per injection (0.3 mL of 200 mg/mL) for the remaining time period by his medical providers.

Data collection

Baseline data were collected three times in the 2 weeks prior to testosterone therapy onset. Data were then collected approximately every 2 weeks thereafter for 1 year. Data included videos of laryngeal structure and function collected via flexible laryngoscopic imaging, acoustic recordings, perceptual measures, and aerodynamic measures collected using clinically-available tools.²¹ Data were collected by study staff who were not blinded to the purpose of the study, but who did not have particular hypotheses about each measure.

Flexible laryngoscopy

Laryngoscopic video and acoustic signals were recorded with the Digital Stroboscopy System (Kay Elemetrics, Lincoln Park, NJ) with both halogen and strobe light source via a distal imaging chip (light source and video processor EPK-1000; pediatric endoscope, VNL-1070STK, 3.3 mm width; both Pentax, Tokyo, Japan). Video was digitized at 30 frames/second with a frame size of 480 × 360 pixels. Tasks are delineated in the appendix and included sustained /i/ at comfortable, high- and low-pitch, and at normal, soft, and loud loudness; pitch and loudness glides; vocal diadochokinesis maneuvers; and singing.

Acoustic recordings

Acoustic recordings generally occurred on the same day as the flexible laryngoscopy recordings and took place in sound-treated rooms (standard acoustic booths produced by IAC Acoustics). Acoustic recordings were made with a standard headset microphone (WH20; Shure, Niles, IL) placed approximately 6–10 cm from the mouth at a 45°

TABLE 1.
Results of Longitudinal Studies in Trans Men Before and After Testosterone Therapy

Study	N	Task	Speaking f_0 Change	Steepest Change	Other Measures
Van Borsel ⁷	2	/a/ and Reading	200–220 Hz to 130–160 Hz	4 mo	Jitter = Shimmer = PFR =*
Damrose ¹²	1	/a/	228 to 113 Hz	3–4 mo	Shimmer ^ NHR ^ PFR =*
Deuster ¹³	11	Reading	79.4 Hz (8.78 ST)	2–3 mo	PFR = Jitter v Shimmer v NHR ~ PFR =
Nygren ⁵	50	Reading	192 to 125 Hz		
Irwig ¹⁴ and Hancock ¹⁶	7	/a/, Reading	135–183 to 87–128 Hz (mean: -6.4 ST)		
Zimman ¹⁷	3	Reading	167–177 Hz to 122–132 Hz		First three formants (F1–3)~

* Authors suggest that PFR reduced, but recalculating results in ST rather than Hz shows a stable PFR (<10% change)

Key =: (average participant shows) no change; ^: increased; v: decreased; ~: inconsistent between participants

Abbreviations: PFR, phonatory frequency range (maximum and minimum f_0 that the speaker can produce); NHR, noise-to-harmonic ratio.

angle from the midline. Signals were preamplified by an RME Quadmic II (RME, Haimhausen, Germany) and sampled at 44,100 Hz with 16-bit resolution using a MOTU UltraLite-mk3 Hybrid (model UltraLite3Hy; MOTU, Cambridge, MA). Recordings were made using SONAR software (Cakewalk, Boston, MA). Tasks are delineated in the appendix and included producing isolated vowels, spontaneous speech, reading, and singing.

Aerodynamic recordings

Aerodynamic recordings were completed using the Phonatory Aerodynamic System (PAS; KayPentax, Lincoln Park, NJ). Tasks are delineated in the appendix and included /a-pa-pa-pa-pa-pa/ and /i-pi-pi-pi/ trains for intraoral subglottal pressure estimates and /pa-pa-pa-pa/ trains for phonation threshold pressure (PTP) estimates. For the subglottal pressure estimates, the participant was instructed to produce /pa/ tokens at a comfortable pitch and loudness at a rate of approximately 90 bpm (demonstrated with metronome during first session). For the PTP measures, the participant was instructed to begin at a comfortable pitch and loudness and decrease his vocal volume until he was unable to phonate.

Habitual recordings

Habitual f_0 recordings were made using a hand-held recorder (H4n Handy Recorder, Zoom, Hauppauge, NY) and a neck-surface mounted accelerometer (Hot Spot accelerometer; K&K Sound, Coos Bay, OR). Accelerometer data were recorded at 44,100 Hz. The participant applied the accelerometer with double-sided tape and carried the recorder in his pocket, recording his voice use for 3–4 hours during the work day. Recordings were made approximately monthly; the participant chose when to apply the sensor and was instructed to go about his day as normal. He did not log his activities separately from the acoustic recordings themselves.

Other data

The participant recorded his testosterone dosage schedule, his testosterone blood levels (only assessed when medically advised, so there is no baseline testosterone level prior to HRT), and his subjective impressions of his voice quality and ease of use, during both daily life and when singing in a choral setting. At each acoustic recording session, the participant recorded his self-rating of difficulty and fatigue of singing in a high-pitched, whispered voice, on a range of 1 (easy, no fatigue) to 10 (most difficult, highest fatigue). The participant also completed the Voice-Related Quality of Life (V-RQOL) scale at each acoustic recording session.

Listener perception of gender

Eight young adult listeners (4 cisgender females, 4 cisgender males; M: 20.9 years, SD: 2.8 years) provided ratings of the participant's gender during a single visit. All listeners were

native speakers of American English, passed a hearing screening at a minimum threshold of 25 dB at octaves from 125 to 8000 Hz, and reported no history of speech, language, or hearing disorders. No listener had any known prior interactions with the primary study participant. Listeners provided written consent in accordance with the Boston University Institutional Review Board.¹

The listener perception study was conducted in a quiet room at the study site. Recordings were presented via circumaural headphones (Sennheiser HD280 Pro). Listeners were first presented with two sample recordings from speakers other than the study participant (one cisgender female, one cisgender male), which they used to adjust the computer volume to a comfortable level. Presentation order of sample recordings was counter-balanced across listeners.

Experimental recordings were then presented to listeners, who were instructed to indicate the speaker's gender and their confidence in that decision by sliding a marker along a 100 mm visual analog scale (VAS). Listeners were told that some recordings may sound alike, but that each recording should be rated separately and carefully. Listeners based their gender perception ratings on excerpts from the *Rainbow Passage* (sentences 2–4) read by the study participant at each timepoint over the course of the study. Excerpts were approximately 13 seconds long (M: 13.33 seconds, range: 12.6–14.6 seconds) and normalized for peak intensity using MATLAB. Listeners rated each sample twice; they were first presented with all 26 recordings in a random order followed by a second presentation of the 26 recordings in a different random order, for a total of 52 recordings. Listeners played each sample once and recorded their ratings on the VAS using a custom-designed interface developed in MATLAB. The VAS ranged from “definitely male” (0 mm) to “definitely female” (100 mm) with intermediate anchors of “probably male/female” (27 mm/73 mm) and “guessing male/female” (49 mm/51 mm).¹¹ Listeners could place the marker at any point along the VAS except the midpoint (50 mm) so that a choice between male and female genders was required. The listening task lasted approximately 20 minutes.

Data analysis

All data analyses, source (ie, acoustic, laryngoscopy video, etc), and the speech or voice task over which the analysis was performed are summarized in Table 2. Details are below. Subjective measures are reported with reliability measures. Acoustic and aerodynamic measures were not reanalyzed for reliability, given that they are objective measures. When data analysts could be blinded, they were. In other cases, analysts were study staff and thus were not

¹While the main study participant did actively consent to all procedures herein, his written consent was not required by the BU IRB, as a case study does not fall under the official definition of research.

¹¹Throughout this study, the terms “male” and “female” are used not as categories of biological sex, but as descriptors of gender. This usage is consistent with norms present in both colloquial speech and existing literature on voice and gender perception.

TABLE 2.
Overview of Measures

	Measure (units)	Source	Token
Pitch and loudness	f_o – min/max/mean (Hz, ST)	Acoustic	Sentences 2–4 of <i>Rainbow Passage</i>
	f_o – habitual (Hz)	Accelerometer	Daily conversation
	f_o variability (Hz, ST)	Acoustic	Sentences 2–4 of <i>Rainbow Passage</i>
	Loudness and variability (dB SPL)	Acoustic	Sentences 2–4 of <i>Rainbow Passage</i>
Physiological measures	Formants (Hz) and laryngeal height (cm)	Acoustic	Steady-state /a/, /i/, /æ/, /eɪ/
	Subglottal pressure estimates (cmH ₂ O)	Aerodynamic	a-pa-pa-pa-pa-pa, i-pi-pi-pi-pi-pi
	Phonation threshold pressure (cmH ₂ O)	Aerodynamic	pa-pa-pa-pa
	Vibratory ratings	Laryngoscopy video	Comfortable modal sustained vowel
Voice quality	Voice quality	Expert listener ratings (CAPE-V)	CAPE-V sentences
	CPP (dB)	Acoustic	Sentences 2–4 of <i>Rainbow Passage</i>
	LH ratio (dB)	Acoustic	Sentences 2–4 of <i>Rainbow Passage</i>
	Airflow (L/s)	Aerodynamic	/a/ during a-pa-pa-pa-pa-pa
	Jitter (%)	Acoustic	Steady-state /a/
	Shimmer (%)	Acoustic	Steady-state /a/
	Harmonics-to-noise ratio (dB)	Acoustic	Steady-state /a/
Perception of gender	Gender perceptual rating	Listener ratings	Sentences 2–4 of <i>Rainbow Passage</i>

blinded. In all cases, there were no specific hypotheses about the amount or direction of change (or stability).

Pitch and loudness correlates

All f_o traces were exported from Praat²² after manual inspection and adjustments were made to Praat settings as needed. Custom scripts in MATLAB (Mathworks, Natick, MA) were used to convert f_o traces to semitones (ST) for further processing. Because of the logarithmic relationship between f_o in Hz and the perception of pitch, measures of differences in f_o (eg, standard deviation as a correlate of intonation, PFR) in Hz will be misleading. This issue is resolved by converting all frequencies to ST as in Equation 1, which represent each frequency as a change in frequency (f) from a reference frequency (f_{ref}).

$$ST = 12 \times \log_2 \left(\frac{f}{f_{ref}} \right) \quad (1)$$

Mean f_o (correlate of pitch) and f_o variability (correlate of intonation). Mean f_o was calculated in Hz and in ST, using the mean f_o from the first baseline session (190.1 Hz) as f_{ref} . For measuring variability, f_o traces were converted to ST via the formula in Equation 1, in which f is the extracted frequency in Hz and f_{ref} is the mean f_o of that session. Variability is then calculated as the standard deviation of f_o traces in ST.

PFR and pitch breaks. PFR is calculated from the minimum modal frequency and maximum falsetto frequency that an individual can produce.²³ The average adult PFR is 38 ST for cisgender men and 37 ST for cisgender women²³ and is used widely in the voice clinic.²⁴ PFR was calculated in ST using Equation 1, with the maximum f_o as f and the minimum f_o as f_{ref} ; maximum and minimum f_o were extracted in Praat from a glissando or discrete half-steps up and down the scale, whichever produced more extreme f_o .²⁵

Pitch breaks during a descending glissando on /i/ during the laryngoscopy examinations were hand-notated during vibratory analysis (see 2.3.2.4 *Vibratory ratings* for more details). The time-locked acoustic signal was examined in Praat and the last f_o before the break and the first f_o after it were extracted. These are used as a marker of stability and control (ie, as a trained singer, the participant should have no pitch breaks during a glissando) and as a marker of the location of the register shift from falsetto to modal.

Habitual f_o . Accelerometer recordings were filtered between 50 and 450 Hz; pitch contours were extracted from Praat manually and evaluated using custom scripts in MATLAB.

Sound pressure level. Sound pressure level (SPL) was calculated as the amplitude of the microphone signal normalized by a calibration procedure. During each recording session, an electrolarynx was placed at the lips and a sound

level meter (CM-150; Galaxy Audio, Wichita, KS) was placed at the microphone. The root-mean-square (RMS) of the microphone amplitude during three different electrolarynx levels were regressed against the known dB SPL measured with the sound level meter.²⁶ The resulting scaling factor and intercept were used to convert the RMS of the microphone signal during the *Rainbow Passage* to dB SPL. The RMS of the microphone amplitude was calculated over 50 ms windows and averaged across the entire sample.

Physiological changes

Formants and estimated vocal tract length. Formants were extracted from prolonged productions of vowels /a, i, æ, eɪ/ by a trained technician, blinded to study hypotheses, using Praat. Formants were calculated over a steady portion of the vowel for a total of 12–20 estimates per timepoint. Estimates were averaged within vowel type and date (eg, all repetitions of /i/ were averaged together) and then across all four vowel types. The fourth formant was used to estimate vocal tract length in cm using Equation 2, in which n is the formant number (4), c is the speed of sound in air (34,300 cm/s), and F_n is the measured formant location in Hz (approximating the fourth vocal tract resonance).

$$\text{vocal tract length} = \frac{(2n-1) * c}{4 * F_n} \quad (2)$$

Subglottal pressure estimates. Raw data were extracted from PAS software and processed further using custom MATLAB scripts. A semi-automated algorithm identified the point of maximum intraoral pressure of the /p/ productions preceding each vowel (ie, /pa/); the peak pressures were each inspected by a trained technician who was blinded to the study hypotheses. Only the middle three productions in each string are used in order to avoid known effects of initial and final utterance positions on intraoral subglottal pressure estimates.²⁷ Estimates were averaged over each utterance to get an average maximum subglottal pressure per timepoint.

PTP. Raw data were extracted from PAS software and processed further using custom MATLAB scripts. These scripts allowed a trained technician, blinded to study hypotheses, to identify the final two peaks of pressure of the /p/ production in each /pa-pa-pa/ train (ie, pressure during the bolded /p/ here: /papapapapa/). The script extracted the maximum pressure during the final two peaks of each /pa-pa-pa/ train and averaged them. PTP for each timepoint was calculated as the average of that mean value over all /pa-pa-pa/ trains produced.

Vibratory ratings. Ratings of the laryngoscopy videos were made by a certified speech-language pathologist specializing in voice using the Voice-Vibratory Assessment With Laryngeal Imaging (VALI) form,²⁸ which elicits ratings of a variety of factors including glottal closure, amplitude of vocal

fold movement, magnitude of mucosal wave, phase closure, and regularity. All ratings were made based on video of comfortable sustained modal phonation. The rater was study staff and as such was not blinded to the purpose of the study; however, the videos were presented in pseudorandomized order so that the timepoint of each video was unknown to the rater.

Voice quality measures

Voice quality changes were assessed perceptually and via acoustic and aerodynamic correlates.

Expert listener ratings. A certified speech-language pathologist specializing in voice completed ratings via the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V)²⁹ as implemented in a custom MATLAB graphical user interface. The speech-language pathologist listened to amplitude-normalized acoustic recordings of all six CAPE-V sentences and provided ratings for overall severity, roughness, breathiness, and strain. Samples were presented in pseudorandomized order, and twenty percent of the samples were repeated for intrarater reliability assessment. The rater was study staff and as such was not blinded to the purpose of the study.

Acoustic measures. Acoustic measures of voice quality were quantified using Analysis of Dysphonia in Speech and Voice software (ADSV)³⁰ or Praat.²² ADSV was used to calculate cepstral peak prominence (CPP) and low-high spectral ratio (LH ratio). Samples were sentences 2–4 of the *Rainbow Passage* and thus match the perceptual study (see 2.3.4 *Listener perception of gender*). Periods of silence between words were removed from each sample using a custom MATLAB script to obtain the most accurate values from the ADSV analysis. Analysis was conducted using the *Rainbow Passage* profile with the cepstral peak extraction range set to encompass the participant's f_o range during connected speech (90–300 Hz).

Praat's "voice report" function was used to estimate jitter, shimmer, and harmonics-to-noise ratio (HNR). Although these measures are no longer preferred as measurements of voice quality,²¹ they are reported in the previous transmasculine literature and thus are presented here. A center portion of each vowel was identified by a trained technician and used for both formant calculation and voice quality measures (see 2.3.2.1 *Formants and estimated vocal tract length*). These were automatically extracted using custom MATLAB scripts to save extracted vowels as separate files and run a voice report using a custom Praat script. Although Praat utilizes several different algorithms to calculate jitter and shimmer, we report jitter ($ppq5$, %) and shimmer ($apq5$, %).

Airflow during vowels. Trains of /a-pa-pa-pa/ were recorded and visualized using PAS software. Oral airflow (L/sec) was measured by manually selecting a stable portion

in the center of each vowel. Only the middle three productions in each string were used to avoid possible effects of initial and final utterance positions. Airflow measurements were not taken from any production of /pa/ for which airflow did not return to 0 L/second between syllables; four such productions were excluded. Airflow was averaged across productions for each timepoint.

Listener perception of gender

Intrarater reliability was assessed via repetition of each voice sample during the listener perception study. Listeners thus rated each sample twice, and Pearson's correlation coefficients were calculated for each listener. Interrater reliability was assessed by calculating the intraclass correlation coefficient (C,k) for consistency.³¹ All listener ratings were averaged to generate a single gender rating of the primary participant's voice at each timepoint.

Statistical approach

Given both the large number of measures and the limitations inherent in a single-subject case study, we present and discuss full time courses of only a few measures of interest (f_o of reading and habitual use, self-ratings of voice issues, and listener perception of gender). The remaining 30 measures are presented in full in the supplemental material for this paper. For each measure, we also performed two sample t tests between baseline and final sessions (three baseline sessions and three final timepoints) to determine significant changes, as well as effect sizes (Cohen's d) to indicate the magnitude of the changes. The P values are all uncorrected, as we wish to provide a summary statistic of which measures have likely changed given that all/many may be expected to change; that is, we have prioritized type I errors over type II errors. This is a conservative position in this case, such that if we conclude that something has not changed, it has likely not changed permanently. Although we are prioritizing long-term (permanent) changes with this data analysis method, all short-term changes (eg, f_o range instability) can be discerned from the graphs in the supplemental material and are discussed individually.

RESULTS

Means and ranges are described below for all measures in detail, including notes of significant differences. Table 3 shows the mean and standard deviation of all measures over the three baseline sessions and the final three sessions, as well as effect sizes (Cohen's d) and P values for changes between these baseline and final session averages (two-sample t tests between three baseline timepoints and final three timepoints).

Medical reports

The participant's total testosterone levels were measured via blood test as medically indicated and are as follows: 408 ng/dL at week 13; 506 ng/dL at week 27; 413 ng/dL at

week 41; 622 ng/dL at week 57. Typical adult cisgender male ranges are 250–827 ng/dL as per the clinical reference for the given immunoassay, while typical adult cisgender female ranges are undetectable. During cisgender male puberty, levels increase gradually from undetectable before the onset of puberty and then from 19 ng/dL to 482 ng/dL.³²

Voice and singing self-report

The participant sang with a high-intensity choir that performs weekly sacred repertoire for a radio audience and a concert series of masterworks. He continued singing throughout this period of voice change (at least 6 hours/week). He moved from second alto to tenor before voice change onset, continued as a tenor throughout the period of this study, and then sang as a baritone/bass as it became clear that both tenor parts sat unnaturally in his changed voice. Comfortable singing ranges changed from F3–D5 to G2–E4. PFR changes are discussed below. His V-RQOL scores did not waver from least-impacted scores (10/50 raw or 100/100 adjusted), indicating that, despite voice changes, he did not note any impact on his quality of life.

Pitch, pitch variability, and loudness

Speaking (reading and habitual)

During the three baseline sessions, the speaker's mean f_o during reading was 183 Hz. During the final three sessions, his mean f_o was 134 Hz. Mean reading f_o thus changed by -5.4 ST. When tracking his habitual pitch, however, the participant's f_o was 211 Hz (SD: 53 Hz) during a baseline measure and 137 Hz (SD: 28 Hz) during the final habitual data collection session, suggesting a change of -7.5 ST. Both of these are significant changes with large effect sizes.

The course of f_o changes for both reading and habitual speaking tasks are shown in Figure 1A. Means and effect sizes between baseline (three baselines) and final (final three recordings) sessions are shown in Table 3. When binning the recordings into 2-month groupings, the largest changes were seen between months 1–2 and 3–4 (-1.8 ST) and months 3–4 and 5–6 (-1.9 ST).

Speaking f_o variability

In the first three baselines, the average speaking variability was 21.3 Hz; by the final three sessions, the variability was 15.9 Hz, which at first suggests a significant change. However, these apparent differences are an effect of the mean f_o changing. When measured as the standard deviation in ST, speaking variability was 1.94 ST in baseline sessions and 2.01 ST in final sessions and not significantly different.

PFR

Figure 2 shows the time course of maximum and minimum f_o changes. In addition, pitch breaks during a descending glissando on /i/ are noted in vertical black lines as an indication of vocal instability and the note(s) of register change.

TABLE 3.
Effect Sizes (Cohen's d) for a Variety of Measures

	Measure	Baseline		Final		Effect Size	P	
		Mean	SD	Mean	SD			
Pitch and loudness	f_o (Hz) - reading	183.49	5.84	134.04	0.83	-11.86	<0.001	
	f_o SD (Hz) - reading	21.34	2.95	15.92	1.44	-2.34	0.046	
	f_o (ST) - reading	-0.74	0.58	-6.17	0.11	-13.03	<0.001	
	f_o SD (ST) - reading	1.94	0.31	2.01	0.20	0.28	0.75	
	Minimum pitch (Hz)	149	8.56	114	5.40	-4.86	0.004	
	Maximum pitch (Hz)	1294	20.9	848	37.65	-14.66	<0.001	
	Phonatory frequency range (ST)	37.46	1.27	34.73	1.42	-2.03	0.07	
	Loudness (dB SPL)	74.64	1.48	75.08	0.80	0.37	0.67	
	Loudness variability (dB SPL)	12.23	2.07	10.41	1.49	-1.01	0.28	
	Physiological	4 th Formant location (Hz)	3694	33.3	3554	13.2	-5.52	0.002
Vocal tract length (cm)		16.3	0.14	16.9	0.06	5.74	0.002	
Subglottal pressure (cmH ₂ O)		8.17	0.98	7.23	0.89	-1.01	0.28	
Phonation threshold pressure (cmH ₂ O)		4.66	0.23	3.37	0.15	-6.60	<0.001	
VALI AP Compression		1.00	0.00	0.67	0.58	-0.82	0.37	
VALI Glottal closure		2.33	0.58	1.67	0.58	-1.15	0.23	
VALI Phase closure		0.80	0.18	1.19	0.10	2.74	0.03	
VALI Regularity		96.67	5.77	100.00	0.00	0.82	0.37	
Voice quality		CAPE-V Overall Severity (%)	3.70	0.99	1.04	0.44	-3.49	0.01
		CAPE-V Roughness (%)	1.23	2.14	1.23	0.33	0	1
	CAPE-V Breathiness (%)	3.32	1.15	0.00	0.00	-4.08	0.007	
	CAPE-V Strain (%)	0.00	0.00	0.00	0.00	0	1	
	CPP (dB) - reading	8.03	0.44	8.75	0.29	1.91	0.08	
	LH (dB)	35.65	0.66	29.38	1.40	-5.71	0.002	
	Airflow (L/s)	0.21	0.01	0.23	0.01	1.64	0.11	
	Jitter (ppq5) %	0.16	0.04	0.22	0.01	2.17	0.06	
	Shimmer (apq5) %	1.46	0.31	3.30	0.35	5.60	0.002	
	Harmonics-to-noise ratio (dB)	23.46	0.88	16.57	0.63	-8.97	<0.001	
Perception of gender rating (100 = definitely female; 0 = definitely male)	96.2	1.24	30.03	2.83	-30.27	<0.001		

Notes: Baseline values are over first three timepoints; final are over final three timepoints. Effect sizes are Cohen's *d* between baseline and final sessions; *P* values are derived from two-sample *t* tests over these six total timepoints (*P* < 0.05 are in bold).

The participant's lowest producible f_o lowered from D#3 to A2 (156 to 109 Hz), and highest f_o changed from E6 to A5 (1318 to 890 Hz), both of which are significant changes with large effect sizes. If measured in Hz, one might interpret these changes as representing a reduction in range. However, when measured in ST to account for the logarithmic relationship between Hz and perceived pitch, the highest note decreased by 6.3 ST and the lower note decreased by 6.2 ST, representing a stable pitch range that shifted down. Accordingly, the average PFR during the baseline sessions was 37.46 ST and 34.73 ST during the final sessions, which was not a significant change.

Speaking amplitude

Loudness was approximated from microphone signal amplitude in dB SPL while the participant read sentences

2–4 of the *Rainbow Passage*. Mean amplitudes were unchanged when comparing the three baselines to the final three sessions: 74.6 dB SPL in baseline and 75.1 dB SPL in the final sessions. Amplitude variability, another correlate of intonation, was 12.2 dB SPL in baseline sessions and ended at 10.4 dB SPL during the final sessions (not significantly different).

Physiological measures

Formants and estimated vocal tract length

Formant location steadily decreased throughout the course of the experiment, resulting in an increasing estimated vocal tract length. The fourth formant had a mean value of 3693 Hz during the first three sessions and 3554 Hz during the final three sessions. These correspond to an estimated vocal tract length of 16.26 cm in the first sessions

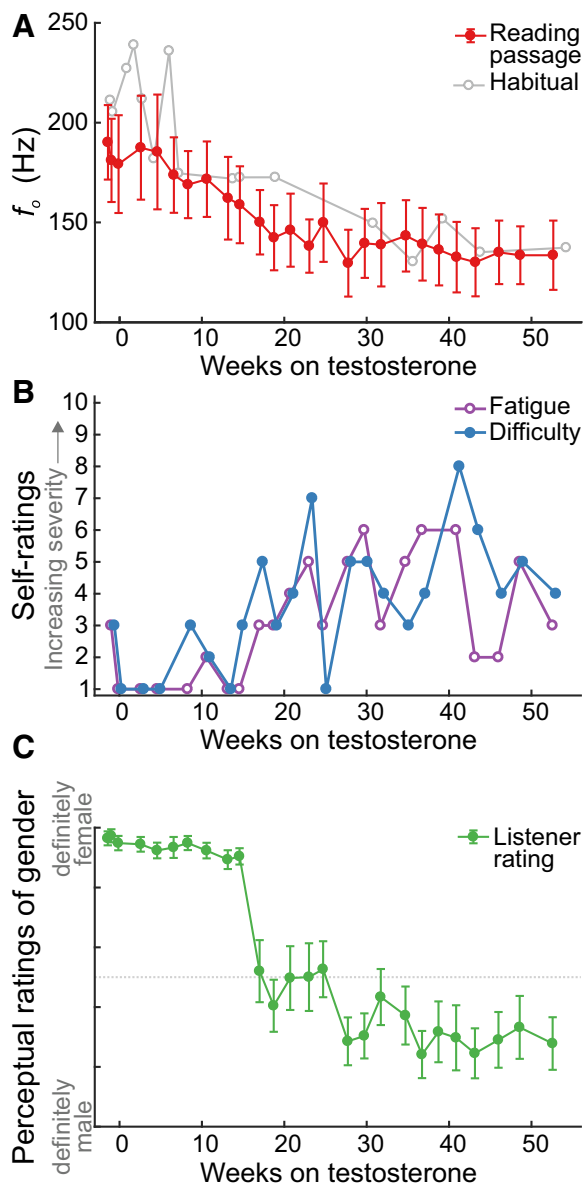


FIGURE 1. Panel A. Speaking f_0 (Hz) during a reading task (red circles) and during habitual speaking throughout the workday (grey open circles). Error bars are standard deviation. Panel B. Participant ratings of fatigue (purple, open circles) and difficulty (blue, closed circles). Panel C. Listener ratings of gender (red, closed circles), from 0 (definitely male) to 100 (definitely female), averaged across listeners. Error bars are standard error. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and 16.90 cm in the final three sessions, both of which are significant changes with large effect sizes.

Subglottal pressure estimates

Intraoral estimates of subglottal pressure suggested that it remained within normal ranges (5.40 cmH₂O [SD 1.37]) for cisgender women and 6.65 cmH₂O [SD 1.98] for cisgender men^{33,34}) throughout the year. The mean subglottal pressure

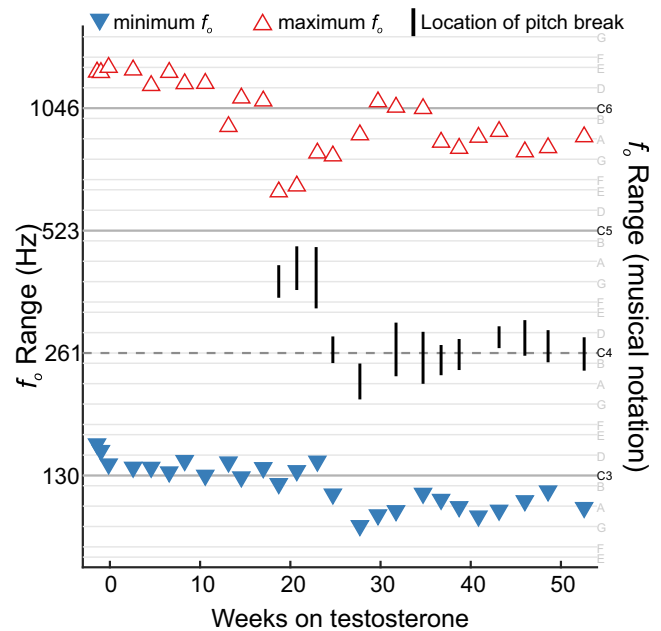


FIGURE 2. Maximum (open red triangles) and minimum (closed blue triangles) singing f_0 . Pitch breaks occurring during a descending glissando, indicative of a change from falsetto to modal, are indicated in vertical black bars. Frequency in Hz of octave Cs are noted on left, with middle C marked with dotted grey line. Musical notes marked with light grey lines with right labels. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

was 8.17 during baseline sessions and 7.23 during the final sessions, which was not a significant change.

PTP

PTP measures gradually declined throughout the year from an average of 4.66 during the three baselines to 3.37 in the final three sessions. This change was significant with a large effect size.

Vibratory ratings

All vibratory ratings from the VALI remained consistent throughout the course the experiment,^{III} with the exception of free edge contour. The rater (as had the SLP performing the flexible laryngoscopy) noted that there was a granuloma on the right vocal process for weeks 34–44 (see Figure 3). Follow-up with an otolaryngologist suggested that this finding was an intubation granuloma acquired when the participant underwent top surgery during week 32. As it was posterior to the vocal process and the participant typically had a posterior gap, this lesion likely did not affect his voice. The participant noted no differences in his voice during that time.

^{III}Although there are no published clinical norms for the VALI, it is validated and the rating clinician indicated that all ratings were within clinically acceptable ranges and thus not indicative of hyperfunction or other voice issues.

Glottal closure was typically rated with a small posterior gap or very small posterior gap (23/26 timepoints), with a rating of “posterior gap” at the first timepoint and “complete closure” at weeks 31 and 34. Anteroposterior compression remained low, between 0–1 out of 5 for all but weeks 4 and 29, which were rated two. Mediolateral constriction was 0 at all timepoints except for the third baseline, which was rated as a one. Regularity was rated 90–100% for all timepoints except weeks 6 and 12, which were rated at 70%.

The ratio of closed phase to open phase varied somewhat over the course of the experiment, with a mean ratio of 0.8:1 during the first three sessions and 1.19:1 during the final three sessions. The maximum ratio was 2:1 during week 27. The minimum was 0.66:1 during the baseline. This measure could not be calculated during week 12 because the strobe system during laryngoscopy did not reach a steady-state to track cycles. Phase symmetry was 100% across all timepoints except week 12, again because the stroboscopy did not track. Mucosal wave was rated an 8/10 at all timepoints. Amplitude was rated as a four for all timepoints except for week 20, which was rated a two.

Self-assessment of fatigue and difficulty

Figure 1B shows the participant's ratings of fatigue and difficulty while using his voice during the biweekly recordings. His ratings were initially low (1–3) in the baseline sessions and for several timepoints thereafter, eventually increasing to steady 4–8 ratings for the bulk of the experimental period, before returning to lower ratings (3–5) for the final sessions.

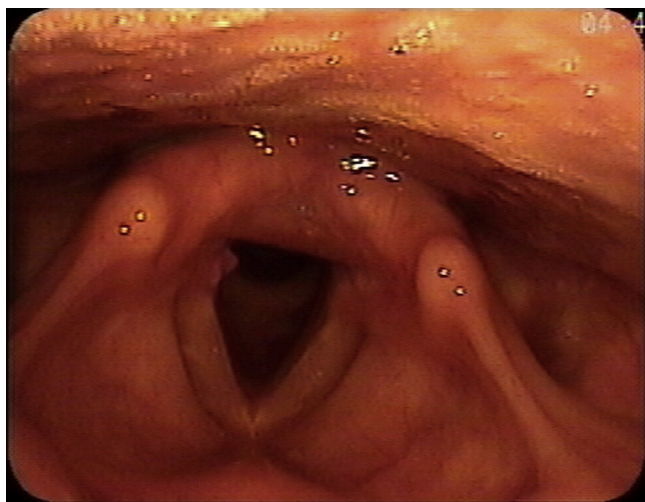


FIGURE 3. View of the intubation granuloma on the right vocal process. Image is from week 36, 1 month following (unrelated) surgery. Granuloma was no longer visible during week 46. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Voice quality

Expert perception of voice quality

All voice quality measures from the CAPE-V remained very low (indicating normal voice quality), with a maximum score of 5.1/100 for any percept by both the main rater and one additional rater who rated 20% of the samples for reliability. Statistically, some changes were significant. Overall severity reduced significantly from 3.7 to 1.04 and breathiness reduced from 3.32 over the three baselines and 0.0 over the final three sessions.

Acoustic and aerodynamic measures of voice quality

CPP increased from the start to end of the experimental period, with an average measure of 8.03 dB in the first three sessions and 8.75 dB in the final three sessions; these differences were not statistically significant and indicate a nondysphonic voice quality throughout.³⁵ CPP ranged from a minimum of 7.51 dB (first baseline) to a maximum of 10.11 dB (week 43), with an overall mean of 8.81 dB.

LH ratio steadily decreased throughout the course of the experiment. The mean LH value was 35.6 dB over the three baselines and 29.4 dB over the final three sessions, a significant change. LH ranged from 28.0 dB (week 48) to 36.1 dB (second baseline), with an overall mean of 31.6 dB; these are all within normal, nondysphonic ranges (within $M \pm 1$ SD from Lowell *et al*³⁵).

Airflow over the vowel /a/ in a /a-pa-pa-pa/ train had a mean of 0.21 L/second during the baselines (SD: 0.01) and 0.23 (SD: 0.01) during the final sessions, values that are within normal ranges and did not change significantly.²⁷ Measures varied throughout the course of the year, with a range of 0.19–0.32 L/second and overall mean of 0.24 L/second.

Three common measures of voice stability and health have been previously reported as aberrant in the literature on trans men. Jitter, representing cycle-to-cycle variability in f_0 , was 0.16% during baseline sessions and 0.22% during the final sessions, which was not a significant change. Shimmer also increased from 1.49% to 3.30% from the baseline to final sessions, and HNR decreased from 23.4 dB to 16.6 dB, both representing significant changes.

Listener perception of gender

Intrarater reliability ranged from 0.53 to 0.95 ($M = 0.82$) for all listeners, and the intraclass correlation coefficient for interrater reliability was 0.96. Thus all listeners were considered sufficiently reliable and consistency across listeners was high. Figure 1C shows the listener ratings of the participant's gender. The mean gender perceptual rating of the participant's voice was 96.2 (SD: 6.3; 0 = definitely male; 100 = definitely female) during the baseline sessions and 30.0 (SD: 26.0) during the final sessions. The participant was reliably identified as female (≥ 65) through the first 15 weeks of testosterone therapy and reliably identified as male (≤ 35) after 37 weeks on testosterone. The mean gender perceptual rating between these timepoints was 42.7 (SD: 9.2),

suggesting some ambiguity in the participant's vocal presentation during this period. Listener perception of the participant's gender showed the largest changes between weeks 15 and 17 (38.4) and weeks 25 and 28 (24.1).

DISCUSSION

The purpose of this study was to comprehensively track one speaker's voice during his first year of testosterone therapy for the purpose of masculinization.

Comparisons to previous transmasculine literature

Overall f_o and the time course of the changes (see full details in supplemental material) observed in this study were consistent with the existing literature. The participant's mean f_o during reading ended at 134 Hz, resulting in a change of -5.4 ST. This change is as expected based on final f_o values from the literature (shown in Table 1): Van Borsel⁷ (130–160 Hz), Nygren⁵ (125 Hz), Irwig¹⁴ (87–128 Hz), and Deuster¹³ (96–140 Hz). This f_o change is also consistent with reference values for a typical cisgender male range (85–155 Hz)³⁶; however, some papers have suggested a more stringent cutoff of 131 Hz.^{2,6} Interestingly, the participant's final speaking f_o and modal singing range are nearly identical to his cisgender brother;^{IV} although all conclusions here must by nature be speculative, this finding suggests that changes may be driven by genetics and that future research may thus be able to predict what changes a transmasculine speaker could expect under testosterone. Further, while the participant's voice changed perhaps moderately compared to other reports (-5.4 ST) and below the 8–12 ST³⁷ change during cisgender male puberty, this degree of change may have been as expected for this speaker, rather than indicating a failure to drop further due to changes in laryngeal structure^{38,39} due to his age upon starting testosterone. Research does in fact suggest that the mechanism of f_o -lowering induced by exogenous testosterone therapy is not identical to that which occurs during typical cisgender male puberty; specifically, the vocal folds are thought to increase mainly in mass rather than in mass and length, because the laryngeal cartilage grows minimally or not at all. This may help to explain the magnitude of f_o change here.⁴⁰

Vocal tract length

Cisgender men typically have lower formants due to body size (longer vocal tracts correspond with lower formants) and due to a lowering of the larynx during puberty, further lengthening the vocal tract. The average adult vocal tract length is 16.9 cm for cisgender men and 14.1 cm for cisgender women, with the large discrepancy emerging after approximately age 15 due to an increase in pharynx length rather than an increase in oral cavity length.⁴¹ The participant's lowered formants suggest his vocal tract length increased to be consistent with mean average value (change

from 16.26 cm to 16.90 cm). Studies of speech often focus on the first three formants (F1, F2, F3),¹⁷ as these can be volitionally modulated by tongue position (eg, the strategy in voice feminization of “fronting” all sounds).⁴² However, higher magnitude formants, such as F4, examined here, are not as affected by articulation, and instead are more influenced by solely the length of the vocal tract from the glottis to the lips.⁴³ To the listener, then, the lowered F4 values may be perceived as suggesting a more masculine voice regardless of F1–F3. Physiologically, these lower F4 values may have been caused by the participant's larynx tilting and lowering on testosterone; functionally, this change could also be caused by the participant speaking with a consistently lowered laryngeal posture. Regardless, these results suggest that the participant's formant resonance (F4) approximated those of cisgender men.

Voice quality

Previous reports have noted detrimental changes in voice quality following testosterone therapy in trans men. The results of our examinations of voice quality are not straightforward, but examination of the full traces (supplemental material) may provide a clearer picture.

Across the entire time period, LH ratio, jitter, shimmer, and HNR appeared to go toward a worsened voice quality^V and stabilize after week 20; of those, jitter had nonsignificant changes from baseline to final and the rest were significantly worse (Table 3). CPP was variable throughout without a clear pattern, and did not significantly change from baseline to final sessions. Airflow increased during weeks 25–40, but appeared to return to baseline levels (nonsignificant change). Of these quantitative measures, only CPP is recommended as an acoustic marker for the assessment of voice quality.²¹

Expert listener perception noted a reduction in overall severity and breathiness, and all measures remained very low throughout (ie, indicating a typical voice; maximally 5.1/100). Perceptual ratings of breathiness significantly decreased (Table 3), indicating *improved* voice quality. Although we did not elicit self-perceived voice quality ratings, the participant did complete the V-RQOL, which remained at the least-impacted level, suggesting that any voice changes did not impact his day-to-day functioning. As a whole, it is unclear whether voice quality changed, and it appears to have not changed in a way that was perceptually meaningful.

Neither the participant nor the expert clinician noted a perception of “entrapment,” a voice quality change thought to be a consequence of exogenous testosterone therapy in transmasculine individuals and/or cisgender women, in which the vocal folds are thought to increase in mass but not in length and be trapped in a restrictive, ossified larynx.^{15,40,44} Perceptually, entrapment has been said to result in a voice that is weak, permanently hoarse, and devoid of

^{IV} *Rainbow Passage* sentences 2–4: 130 Hz (SD: 26.0 Hz); Phonatory frequency range: 100 Hz–598 Hz; comfortable modal singing range: A2–D4.

^V “Worse” is indicated by a lower LH ratio, HNR, and CPP, and higher jitter and shimmer.

the “right” harmonics.¹⁵ While our clinical protocols did not include a specific rating of “entrapment,” these percepts are also consistent with those experienced by individuals with hyperfunctional voice disorders, which are assessed with the protocols used here (eg, the CAPE-V ratings). Even so, future research may wish to examine this specific percept in more detail.

It is unknown whether this participant's experience represents a typical transmasculine experience. There have been reports of long-term changes in voice quality, both perceptually and those captured by some acoustic measures (albeit those that are generally no longer recommended), as well as the expected short-term vocal instability. What we can conclude, however, is that it is not inevitable that an individual's voice quality will be negatively impacted. Transmasculine individuals are routinely warned (by medical staff, the trans community, and others) that their voice quality will suffer, which is of concern particularly to singers and other heavy voice users. Further careful research should examine self- and expert-listener perception as well as validated/recommended acoustic measures to determine the likelihood of long-term voice impacts.

Listener perception of gender

Very few studies^{9,10} have assessed how listeners perceive the gender of transmasculine speakers, and none have measured how these perceptions may change as speakers undergo testosterone therapy. The present case demonstrated that listener perception of the participant's gender shifted over the course of the study. The shift in gender perception correlated strongly ($r = 0.908$, $P < 0.001$) with decreases in the participant's f_o . This association differs from the findings of Van Borsel *et al*,¹⁰ who reported no correlation between f_o and listener ratings of transmasculine speakers' “maleness.” Note, however, that these were cross-speaker studies rather than within-speaker longitudinal measures as in the present study; thus the speakers had potentially all crossed the perceptual threshold. The point at which the shift in gender perception occurred in this case also differs from past studies. Here, the participant was perceived as male with an $f_o < 140$ Hz, well below Spencer's⁴⁵ reported 160 Hz ceiling for male gender perception.

The speaker's f_o decreased fairly linearly from weeks 0–30, after which it remained steady (Figure 1). Interestingly, there was a nonlinearity in the perception of the participant's gender between an f_o of 158 Hz and 150 Hz. This finding corresponds well to previous reports of typical cisgender male ranges (85–155 Hz) and typical cisgender female ranges (165–255 Hz).³⁶

A potential factor in these different findings is listener characteristics. In a framework that regards listeners as active participants in gender attribution,⁸ each listener's individual experiences with gender inform their assumptions about a speaker's identity. These experiences are vast, changeable, and infinitely varied across listeners. No attempt was made to characterize the complex experiences

with gender for the listeners in this study beyond collecting gender identity data. These data show that the listener perceptions reported here are from the perspectives of cisgender individuals. Those of individuals of other genders might differ in meaningful ways, as may those of listeners with experience with transmasculine friends or family. There is, of course, heterogeneity among the perceptions of cisgender individuals, given their own experiences with gender and with gender-expansive individuals. For example, while the gender ratings of most of our listeners followed a similar pattern, two listeners stood out for their unique responses. One submitted ratings that were almost all 0 (definitely male) or 100 (definitely female); the other always rated the participant as female but also showed the least intrarater reliability (all ratings ≥ 65 ; reliability = 0.53). Without further information on how these individuals define and interpret gender-based norms of communication, these results are seemingly anomalous, but they do call for further research into the influence of listener characteristics and experiences on gender perception. We also wish to note that we are not attempting to say that naïve cisgender listeners are, or should be, the final arbiters of a speaker's perceived gender. Instead, we wished to capture some aspect of how the participant's voice might be perceived in the world, such as when he makes a phone call; this is a common source of misgendering and thus a concern that is frequently used as a measure of voice satisfaction/concern.^{5,8}

Other factors

Hz versus ST

Some previous reports suggest a reduction in singing range, which is of concern to transmasculine speakers, particularly those who sing. However, our results match those of others, which suggest that, when measured in ST, the participant's PFR shifted down by approximately 6 ST.

Intubation granuloma

The participant underwent surgery around week 32 after beginning testosterone. Following that surgery, the SLP performing the laryngoscopy noted an intubation granuloma (confirmed via an otolaryngologist; see Figure 3). It was posterior to the vocal fold processes and did not appear to interfere with function; thus we considered this lesion an incidental finding, but note that clinicians should recall that transgender individuals are likely to have surgeries periodically and may suffer intubation trauma that should be considered in the case of sudden voice changes.

Chest binding

Transmasculine individuals often wear a restrictive chest binder and may use a hunched posture to mask feminine chest contours; both of these may affect proper respiratory support for voice production.¹ The participant wore a chest binder for recording sessions in the baselines through

week 35, after which he was healed enough from surgery to wear no binder or dressings. For this participant, there were no apparent changes between weeks 35 and 40 that might be directly attributable to the binder. However, this is an important consideration for clinicians and future studies in transmasculine individuals.

Baseline/final sessions evaluation versus intermediate changes

For simplicity and due to the large number of measures and data types examined, we have presented primarily baseline and final session measures and indicated whether those measures changed during the course of testosterone therapy (Table 3 for effect sizes and paired *t* test results). Although Figures 1 and 2 show timecourses for several important measures, the timecourses of all measures may be of interest and are thus shown in the Supplemental materials. Although most measures show a pattern consistent with the *t* test results in Table 3 (ie, those that showed nonsignificant changes remained stationary throughout, and those that showed significant changes had a consistent trajectory throughout), a few outliers may be worth considering. For example, pitch range was consistent from baseline to the final sessions, but had a very reduced range (below 30 ST) during weeks 18–21. Similarly, airflow was unchanged between baseline and the final sessions, but was elevated weeks 27–30. This change preceded the granuloma, and thus may represent some other instability in the vocal motor control system.

Clinical recommendations

Clinical recommendations drawn from a single case study, even when comprehensive and faithful to reporting guidelines,¹⁹ are offered with sufficient caution regarding their generalizability. Nevertheless, the present case supports a departure from common advice to transmasculine speakers regarding the potential effects of testosterone on pitch range. Concerns about a reduction in pitch range were not borne out in this case. When measured in ST, the participant's range remained stable while shifting downward. This result will be of particular interest to transmasculine singers and their clinicians. Further, the different methods of eliciting f_0 (eg, habitual speech during the workday versus elicited tasks of producing isolated vowels, reading, or spontaneous speech under recording conditions) resulted in different mean f_0 values. While those results are not unexpected, this does indicate that a stringent cutoff drawn from the literature is not recommended as a specific goal or measure of acceptably masculine voice. Similarly, pitch and pitch contours deemed masculine and feminine vary, sometimes widely, between cultures. Thus any goals must be set within the client's cultural norms.

Limitations and future directions

This case study involved only one participant. In addition to the limitations inherent in any case study, this particular

participant may not represent a typical transmasculine speaker beginning testosterone. During the course of the study, he worked in a research laboratory focused on voice and voice therapy and sang in a choir with professional singers and voice pedagogues. He was aware of the purpose of the recordings. Although he made no conscious effort to masculinize his voice except via testosterone, these results may not be representative of other transmasculine people. His age is also relevant (30–31), as many transmasculine people are now receiving hormone therapy at younger ages. As speakers age, their laryngeal cartilages gradually ossify.^{38,39} Thus younger speakers may see larger or different changes.

In addition, other acoustic measures have been suggested as markers of masculinity (eg, center of gravity of /s/); we have focused on voice, specifically, but further work on articulatory and phonetic measures is warranted.^{46,47} The participant did not enroll in voice therapy or have voice lessons at any point before or during this research study, although he continued as an active choral singer. Future studies should examine the possible effects of voice therapy either concurrently or following voice lowering on testosterone. Evidence-based practice is vital in this area, especially given the social and economic barriers to voice therapy for transmasculine individuals.¹⁷

This study is an example of participatory action research in which the participant contributed to research design and research question formulation.²⁰ This approach aims to ensure that concerns about outcomes and disparities held by members of the community being studied are not overlooked, as may occur when relying solely on outside researchers' perspectives.¹⁷ As a single-subject case study, however, such contributions reflect the concerns of an individual member of the transmasculine community and so may not be representative of the larger group. Future work on transmasculine voice should consider the participatory action approach as a means of ensuring research is meaningful to the transmasculine community.

CONCLUSIONS

This paper presents a previously-unavailable in-depth assessment of one transmasculine individual's speech and voice changes during the first year of hormone therapy with testosterone. Throughout the course of 1 year, the participant's speaking f_0 reduced from a typical cisgender female range to a typical cisgender male range. His singing range shifted approximately 6 ST, from an alto range to a baritone range. Physiological and perceptual measures indicated a healthy voice production throughout the course of the experiment. Interestingly, acoustic measures indicated that the participant's vocal tract increased in length by 0.6 cm, suggesting that his larynx may have dropped or tilted as is expected in cisgender male puberty. Finally, listeners consistently attributed a male gender to the participant after approximately 37 weeks on testosterone. The data in this case study add to the scant literature

on transmasculine voice. Future studies can assess a standard time-course for changes, which changes relate most to listener perception, and which factors could be targets of speech therapy for speakers who do not find their voices to be satisfactory.

ACKNOWLEDGMENTS

The authors wish to thank Daniel Buckley for his help with laryngoscopy, perceptual evaluation, and discussions about this manuscript. We wish to further thank Jake Noordzij, Zach Morgan, Lauren Maclellan, and Lin Zhang for their help with data processing. Finally, we wish to thank Graham Grail and Carolyn Hodges-Simeon for their discussions about existing transmasculine literature and Jessica Kramer for discussions of participatory action research. This work was supported by the National Institutes of Health – National Institute on Deafness and Other Communication Disorders under grants [F31 DC014872](#) (GJC), [R01 DC015570](#) (CES), [T32 DC013017](#) (Christopher A. Moore), and [F32 DC017637](#) (GJC).

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version, at <https://doi.org/10.1016/j.jvoice.2019.03.006>.

APPENDIX

Laryngoscopy protocol

Light Source	Tasks
Strobe	Sustain /i/ at comfortable pitch/loudness Soft /i/ Loud /i/ High-pitch /i/ Low-pitch /i/ Loudness glides (soft→loud→soft) Vocal diadochokinesis maneuvers: /ihi ihi ihi ihi/, /isi isi isi isi/, /i/-sniff-/i/-sniff-/i/-sniff Sing happy birthday Sing warm-ups with melisma (ie, do-mi-re-fa-mi-so etc, all on /i/) Pitch glides (low→high and high→low) x 2
Halogen	Loudness glides (soft→loud→soft) Vocal diadochokinesis maneuvers: /ihi ihi ihi ihi/, /isi isi isi isi/, /i/-sniff-/i/-sniff-/i/-sniff Sing happy birthday Sing warm-ups with melisma (ie, do-mi-re-fa-mi-so etc, all on /i/) Pitch glides (low→high and high→low) x 2

Acoustic & aerodynamic recordings protocol

Measure	Token	Instructions
Isolated vowels	/a/, /i/, /u/, /æ/, /eɪ/, /ou/	Repeat 5 times each; ~3 s in duration
Spontaneous speech		Suggested prompt: -What did you do last weekend?
Reading	Rainbow passage; CAPE-V sentences; reading passage loaded with corner vowels; SIT sentences	Normal pitch/loudness
Singing	/i/	Habitual pitch Low-pitch High-pitch Low, glide to high High, glide to low Low, scale to high (step) High, scale to low (step) Sing warm-ups with melismas at different ranges (ie, do-mi-re-fa-mi-so etc, all on /i/)
Subglottal pressure estimates	a-pa-pa-pa-pa-pa, i-pi-pi-pi-pi-pi	Record with mask and straw via PAS; simultaneous mic & accel recordings (metronome × 90 bpm-ish) 2 recordings × 8 strings per recording Normal pitch and loudness
Phonation threshold pressure	pa-pa-pa-pa	Record with mask and straw via PAS; simultaneous mic & accel recordings 2 recordings Begin at typical pitch and loudness Decreased vocal volume until unable to phonate any longer *no breaths between /pa/s

REFERENCES

- Davies S, Papp VG, Antoni C. Voice and communication change for gender nonconforming individuals: giving voice to the person inside. *Int J Transgenderism*. 2015;16:117–159. <https://doi.org/10.1080/15532739.2015.1075931>.
- Azul D. Transmasculine people's vocal situations: a critical review of gender-related discourses and empirical data. *Int J Lang Commun Disord*. 2015;50:31–47. <https://doi.org/10.1111/1460-6984.12121>.

3. Zimman L. Transgender voices: insights on identity, embodiment, and the gender of the voice. *Lang Linguist Compass*. 2018;12:e12284. <https://doi.org/10.1111/ln3.12284>.
4. McNeill EJM, Wilson JA, Clark S, et al. Perception of voice in the transgender client. *J Voice*. 2008;22:727–733. <https://doi.org/10.1016/j.jvoice.2006.12.010>.
5. Nygren U, Nordenskjöld A, Arver S, et al. Effects on voice fundamental frequency and satisfaction with voice in trans men during testosterone treatment—a longitudinal study. *J Voice*. 2016;30:766. e23-766.e34.
6. Ziegler A, Henke T, Wiedrick J, et al. Effectiveness of testosterone therapy for masculinizing voice in transgender patients: a meta-analytic review. *Int J Transgenderism*. 2018;19:25–45. <https://doi.org/10.1080/15532739.2017.1411857>.
7. Van Borsel J, De Cuyper G, Rubens R, et al. Voice problems in female-to-male transsexuals. *Int J Lang Commun Disord*. 2000;35:427–442. <https://doi.org/10.1080/136828200410672>.
8. Azul D, Arnold A, Neuschaefer-Rube C. Do transmasculine speakers present with gender-related voice problems? Insights from a participant-centered mixed-methods study. *J Speech Lang Hear Res*. 2018;61:25. https://doi.org/10.1044/2017_JSLHR-S-16-0410.
9. Scheidt D, Kob M, Willmes K, et al. Do we need voice therapy for female-to-male transgenders? 2004 IALP Congress Proceedings. Brisbane, Australia;2004.
10. Van Borsel J, de Pot K, De Cuyper G. Voice and physical appearance in female-to-male transsexuals. *J Voice*. 2009;23:494–497. <https://doi.org/10.1016/j.jvoice.2007.10.018>.
11. Cosyns M, Van Borsel J, Wierckx K, et al. Voice in female-to-male transsexual persons after long-term androgen therapy. *Laryngoscope*. 2014;124:1409–1414. <https://doi.org/10.1002/lary.24480>.
12. Damrose EJ. Quantifying the impact of androgen therapy on the female larynx. *Auris Nasus Larynx*. 2009;36:110–112. <https://doi.org/10.1016/J.ANL.2008.03.002>.
13. Deuster D, Matulat P, Knief A, et al. Voice deepening under testosterone treatment in female-to-male gender dysphoric individuals. *Eur Arch Oto-Rhino-Laryngology*. 2016;273:959–965. <https://doi.org/10.1007/s00405-015-3846-8>.
14. Irwig MS, Childs K, Hancock AB. Effects of testosterone on the transgender male voice. *Andrology*. 2017;5:107–112. <https://doi.org/10.1111/andr.12278>.
15. Constansis AN. The changing female-to-male (FTM) voice. *Radix Musicol* 2008;3.
16. Hancock AB, Childs KD, Irwig MS. Trans male voice in the first year of testosterone therapy: make no assumptions. *J Speech Lang Hear Res*. 2017;60:2472. https://doi.org/10.1044/2017_JSLHR-S-16-0320.
17. Zimman L. *Voices in Transition: Testosterone, Transmasculinity, and the Gendered Voice among Female-to-Male Transgender People*. Dissertation, University of Colorado-Boulder; 2012. https://scholar.colorado.edu/ling_gradetds/15.
18. Azul D, Nygren U, Södersten M, et al. Transmasculine people's voice function: a review of the currently available evidence. *J Voice*. 2017;31:26. 1e9-261.e23. <https://doi.org/10.1016/j.jvoice.2016.05.005>.
19. Gagnier JJ, Kienle G, Altman DG, et al. The CARE guidelines: consensus-based clinical case reporting guideline development. *J Med Case Rep*. 2013;7:223. <https://doi.org/10.1186/1752-1947-7-223>.
20. Minkler M, Wallerstein N, eds. *Community-Based Participatory Research for Health: From Process to Outcomes*. 2nd ed. San Francisco, CA: Jossey-Bass; 2013.
21. Patel RR, Awan SN, Barkmeier-Kraemer J, et al. Recommended protocols for instrumental assessment of voice: american speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. *Am J Speech-Language Pathol* 2018;1. https://doi.org/10.1044/2018_AJSLP-17-0009.
22. Boersma P, Weenink D. Praat: doing phonetics by computer. *Computer program*; 2015. <http://www.fon.hum.uva.nl/praat/>.
23. Hollien H, Dew D, Phillips P. Phonational frequency ranges of adults. *J Speech Lang Hear Res*. 1971;14:755. <https://doi.org/10.1044/jshr.1404.755>.
24. Zraick RI, Nelson JL, Montague JC, Monoson PK. The effect of task on determination of maximum phonational frequency range. *J Voice*. 2000;14:154–160. [https://doi.org/10.1016/S0892-1997\(00\)80022-3](https://doi.org/10.1016/S0892-1997(00)80022-3).
25. Barrett EA, Lam W, Yiu EML. Elicitation of minimum and maximum fundamental frequency and vocal intensity: discrete half steps versus glissando. *J Voice*. 2018;0. <https://doi.org/10.1016/j.jvoice.2018.09.023>.
26. McKenna VS, Stepp CE. The relationship between acoustical and perceptual measures of vocal effort. *J Acoust Soc Am*. 2018;144:1643–1658. <https://doi.org/10.1121/1.5055234>.
27. Holmberg EB, Hillman RE, Perkell JS. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am*. 1988;84:511–529. <https://doi.org/10.1121/1.396829>.
28. Poburka BJ, Patel RR, Bless DM. Voice-vibratory assessment with laryngeal imaging (VALI) form: reliability of rating stroboscopy and high-speed videoendoscopy. *J Voice*. 2017;31:513.e1-513.e14. <https://doi.org/10.1016/J.JVOICE.2016.12.003>.
29. Kempster GB, Gerratt BR, Verdolini Abbott K, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18:124. [https://doi.org/10.1044/1058-0360\(2008\)08-0017](https://doi.org/10.1044/1058-0360(2008)08-0017).
30. Awan SN. *Analysis of dysphonia in speech and voice: an application guide. Computer program manual*. 2011.
31. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996. <https://doi.org/10.1037/1082-989X.1.1.30>.
32. August GP, Grumbach MM, Kaplan SL. Hormonal changes in puberty: III. Correlation of plasma testosterone, LH, FSH, testicular size, and bone age with male pubertal development. *J Clin Endocrinol Metab*. 1972;34:319–326. <https://doi.org/10.1210/jcem-34-2-319>.
33. Zraick RI, Smith-Olinde L, Shotts LL. Adult normative data for the KayPENTAX Phonatory Aerodynamic System Model 6600. *J Voice*. 2012;26:164–176. <https://doi.org/10.1016/J.JVOICE.2011.01.006>.
34. Zraick RI, Smith-Olinde L, Shotts LL. Erratum: “Adult Normative Data for the KayPENTAX Phonatory Aerodynamic System Model 6600.” *Journal of Voice*. 2012;26:164-176. *J Voice*. 2013;27:2–4. <https://doi.org/10.1016/j.jvoice.2011.01.006>.
35. Lowell SY, Colton RH, Kelley RT, et al. Predictive value and discriminant capacity of cepstral- and spectral-based measures during continuous speech. *J Voice*. 2013;27:393–400. <https://doi.org/10.1016/J.JVOICE.2013.02.005>.
36. Fitch JL, Holbrook A. Modal vocal fundamental frequency of young adults. *Arch Otolaryngol - Head Neck Surg*. 1970;92:379–382. <https://doi.org/10.1001/archotol.1970.04310040067012>.
37. Hollien H, Green R, Massey K. Longitudinal research on adolescent voice change in males. *J Acoust Soc Am*. 1994;96:2646–2654. <https://doi.org/10.1121/1.411275>.
38. Hatley W, Evison G, Samuel E. The pattern of ossification in the laryngeal cartilages: a radiological study. *Brit. J. Radiol*. 1965;38.
39. Garvin HM. Ossification of laryngeal structures as indicators of age. *J Forensic Sci*. 2008;53:1023–1027. <https://doi.org/10.1111/j.1556-4029.2008.00793.x>.
40. Romano T. *The singing voice during the first two years of testosterone therapy: working with the trans or gender queer voice*. Dissertation, University of Colorado-Boulder; 2018. https://scholar.colorado.edu/mvoc_gradetds/1/.
41. Goldstein UG. *An articulatory model for the vocal tracts of growing children*. Dissertation, Massachusetts Institute of Technology; 1980. <http://hdl.handle.net/1721.1/16118>.
42. Adler RK, Hirsch S, Pickering J. *Voice and Communication Therapy for the Transgender/Gender Diverse Client: A Comprehensive Clinical Guide*. 3rd ed.; 2018.
43. Lammert AC, Narayanan SS. On short-time estimation of vocal tract length from formant frequencies. *PLoS One*. 2015;10 e0132193. <https://doi.org/10.1371/journal.pone.0132193>.

44. Baker J. A report on alterations to the speaking and singing voices of four women following hormonal therapy with virilizing agents. *J Voice*. 1999;13:496–507. [https://doi.org/10.1016/S0892-1997\(99\)80005-8](https://doi.org/10.1016/S0892-1997(99)80005-8).
45. Spencer LE. Speech characteristics of male-to-female transsexuals: a perceptual and acoustic study. *Folia Phoniatr Logop*. 1988;40:31–42. <https://doi.org/10.1159/000265881>.
46. Zimman L. Gender as stylistic bricolage: transmasculine voices and the relationship between fundamental frequency and /s/. *Lang Soc* 2017. <https://doi.org/10.1017/s0047404517000070>.
47. Zimman L. Variability in /s/ among transgender speakers: evidence for a socially grounded account of gender and sibilants. *Linguistics* 2017. <https://doi.org/10.1515/ling-2017-0018>.