Acoustic Predictors of Gender Attribution, Masculinity −Femininity, and Vocal Naturalness Ratings Amongst Transgender and Cisgender Speakers

*,† Teresa L.D. Hardy, *Jana M. Rieger, ‡,*Kristopher Wells, and *Carol A. Boliek, *†‡Edmonton, Canada

Abstract: Purpose. This study aimed to identify the most salient set of acoustic predictors of (1) gender attribution; (2) perceived masculinity−femininity; and (3) perceived vocal naturalness amongst a group of transgender and cisgender speakers to inform voice and communication feminization training programs. This study used a unique set of acoustic variables and included a third, androgynous, choice for gender attribution ratings.

Method. Data were collected across two phases and involved two separate groups of participants: communicators and raters. In the first phase, audio recordings were captured of communicators (n = 40) during cartoon retell, sustained vowel, and carrier phrase tasks. Acoustic measures were obtained from these recordings. In the second phase, raters (n = 20) provided ratings of gender attribution, perceived masculinity−femininity, and vocal naturalness based on a sample of the cartoon description recording.

Results. Results of a multinomial logistic regression analysis identified mean fundamental frequency (f_o) as the sole acoustic measure that changed the odds of being attributed as a woman or ambiguous in gender rather than as a man. Multiple linear regression analyses identified mean f_o, average formant frequency of /i/, and mean sound pressure level as predictors of masculinity−femininity ratings and mean f_o, average formant frequency, and rate of speech as predictors of vocal naturalness ratings.

Conclusion. The results of this study support the continued targeting of f_o and vocal tract resonance in voice and communication feminization/masculinization training programs and provide preliminary evidence for more emphasis being placed on vocal intensity and rate of speech. Modification of these voice parameters may help clients to achieve a natural-sounding voice that satisfactorily represents their affirmed gender.

Key Words: Transgender−Acoustics−Gender−Femininity−Naturalness−Voice.

INTRODUCTION

The discrepancy between gender identity (ie, internal concept of one's being a man or a woman)[11] and gender attribution (ie, how one's gender is perceived by others)[1] experienced by many transgender women can have negative consequences on quality of life,[2] life participation,[3−5] and even personal safety.[5−7] Consequently, developing a gender expression that is congruent with gender identity can be an important part of the gender role transition process and often motivates transgender women to seek voice and communication feminization training from speech-language pathologists (SLPs). The SLP helps these individuals modify aspects of their communication that influence how others perceive them.[8] Appropriate selection of training targets requires that decisions be informed by research evidence.

Sound of the voice is known to be a salient cue to gender attribution.[9−11] More specifically, speaking fundamental frequency (f_o) and vocal tract resonance (as measured by vowel formant frequencies) appear to be particularly important cues; this holds true for both transgender and cisgender speakers.[11−15] Several other aspects of communication have been proposed as cues to gender attribution such as measures related to intonation, rate of speech (RoS), loudness, and voice quality.[10,12,13,16−18] The relative contribution of each of these variables to gender attribution currently is not well understood, especially when considering transgender women speakers. Understanding the relative contributions of voice and speech variables to gender attribution in cisgender and transgender women speakers is the focus of the present study.

Speaking fundamental frequency

It has been well established in both transgender and cisgender literature that f_o plays an important, perhaps even critical, role in gender attribution.[10,12,14,19−24] In fact, results of a meta-analysis exploring the contribution of f_o to gender attribution showed that this measure accounted for 41.6% of the variance in attributed gender.[1] This voice measure

Accepted for publication October 1, 2018.

From the *Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, Alberta, Canada; †Glenrose Rehabilitation Hospital, Alberta Health Services, Edmonton, Alberta, Canada; and the ‡Department of Child & Youth Care, Faculty of Health & Community Studies, MacEwan University, Edmonton, Alberta, Canada.

Address correspondence and reprint requests to Teresa L.D. Hardy, Faculty of Rehabilitation Medicine, University of Alberta, 3-48 Corbett Hall, Edmonton, AB T6G 2G4, Canada. E-mail: teresa.hardy@ualberta.ca, jana.rieger@ualberta.ca, Kristopher.Wells@Macewan.ca, carol.boliek@ualberta.ca

Journal of Voice, Vol. 34, No. 2, pp. 300.e11−300.e26
0892-1997

https://doi.org/10.1016/j.jvoice.2018.10.002

[1]The terminology used to describe transgender and gender diverse people is evolving. For the purposes of the present manuscript, terminology is differentiated between sex (e.g., male/female), gender (e.g., man, woman, transgender woman), and attributes or characteristics that may overlay those terms (i.e., masculinity−femininity). Although these concepts are related, especially when considering the socially constructed nature of gender roles and expression, they are not synonymous. Gender attribution and masculinity−femininity ratings both were included in the present study to provide a richer understanding of the relationships between communication and gender attribution and to acknowledge the spectrum of possible gender presentations and attributions. Whereas the authors do not wish to promote gender normative ideology, they also recognize that relationships between communication characteristics and attributions made by communication partners exist within a social context. As such, understanding these relationships necessitates acknowledgement of prevailing social belief systems, especially considering the potential adverse impact on personal safety and social participation when incongruences exist.

also has been positively related to femininity ratings made by both unfamiliar listeners and transgender women, themselves.[18,20,25–27] It is important to note, however, that although a good deal of the variance in gender attribution measures was explained by f_o in the meta-analysis, there remained 58.4% to be explained by other communication factors. Furthermore, there is evidence to suggest that the role of these other factors may be more pronounced during connected speech tasks[13] and when f_o falls within a "gender ambiguous" or neutral range,[13,20,23] defined as 150–185 Hz.[28] This same range has been suggested as a gender-acceptable target for communication feminization training.[29] As a result, it is important to identify these other factors and to explore their interaction with f_o in contributing to measures of gender attribution.

Vocal tract resonance and vowel formant frequencies

Vowel formant frequencies have long been suspected to play a key role in distinguishing voices perceived as belonging to men and women[13,14,20,30,31] and now are generally accepted to be contributors to gender attribution.[12] They also have been related to vocal naturalness ratings for transgender women.[27] What remains unclear, however, is the salience of each vowel formant and the strength of the overall contribution.

Studies have shown relationships between auditory-perceptual ratings of gender and/or femininity and each vowel formant (F1, F2, F3) in isolation[13,20,25,30,31] as well as between those ratings and an average of the three formants together.[13] However, significant relationships have only been identified between these variables when measured in the context of connected speech and not isolated vowel productions.[19,27,32,33] Moreover, vowel formant frequencies account for more of the variance in gender attribution ratings when measured in connected speech than in isolated vowels.[13,19] Studies investigating relationships between auditory-perceptual ratings and vowel formant frequencies should, therefore, use connected speech as the capture task.

Despite the compelling evidence for the importance of vowel formant frequencies in contributing to gender attribution and potentially natural-sounding voices, they do not seem to be sufficient for consistent correct gender attribution even when combined with f_o.[14] For that reason, it is important that other acoustic measures be explored as potential contributors as well.

Intonation

Intonation or melody is another aspect of communication that commonly is believed to differ between masculine and feminine communication styles and contribute to natural-sounding speech. Research has shown that transgender women who were attributed as women used more upward intonation shifts, greater frequency variability, and semitone (ST) range, and had a higher upper limit for f_o in connected speech tasks such as reading, picture description, and responding to questions.[17,23,32] Higher upper limit for f_o also was found to significantly positively correlate with femininity ratings.[32] These studies suggest that more feminine communication styles are associated with greater variability in intonation. Other studies, however, have either found the opposite relationship (ie, femininity ratings were significantly related to a narrow ST range)[18] or failed to find relationships between intonation measures and femininity ratings at all.[17]

Taken together, these studies suggest that the contribution of intonation measures to gender attribution is not yet fully understood. Such measures may have some influence on gender attribution, particularly parameters such as ST range and percentage of upward intonation shifts. Further research is needed to confirm previous results and determine the relative salience of intonation as a cue to gender attribution. Research also is needed in other areas related to prosody such as speaking rate.

Rate of speech

The cisgender literature is inconclusive with respect to gender differences in RoS. Some studies suggest that men's speech rate is faster than women's in both reading and conversational contexts,[34,35] whereas others report RoS to be similar between the genders.[36–38]

There are relatively few data with respect to RoS in transgender women speakers and the data that do exist reflect the same inconsistencies seen in the cisgender literature. For example, early studies found transgender women speakers used a slower RoS when reading words and sentences in the feminine versus masculine gender presentation,[30,39] suggesting they believed a slower RoS sounded more feminine. Conversely, Van Borsel and Maesschalck[38] found no differences in RoS between cisgender men, cisgender women, and transgender women speakers when reading a standardized passage.

Of note, none of the studies involving transgender women investigated differences in RoS as a function of gender attribution or masculinity–femininity, which are more relevant to communicative success than gender status. Moreover, previous studies used reading tasks and did not investigate the role of RoS in the more ecologically valid conversational context. Further research is needed to explore RoS as a potential gender marker.

Sound pressure level

According to Gelfer and Young,[40] sound pressure level (SPL) (perceptual correlate = loudness) is one of the most salient perceptual features of the voice and has been shown to differ in small but significant ways between genders in cisgender speakers. For example, in a conversational context, men speak 2–3 dBSPL louder than women[40,41] and have a wider conversational loudness range, which relates to the use of greater sound pressures in their inflections.[40]

Although this parameter has a presence in clinical discussions of gender-related communication differences to be considered in voice and communication feminization training,[12,16,22,42] very little research has investigated the role of SPL in gender attribution, especially as it relates to transgender speakers. The few studies that have been published have found louder voices to be associated with masculine gender presentations[30,39,43] and gender attributions[10] in connected speech tasks. These results were based on a small number of participants (n = 4−6).

Voice quality and vocal perturbation

Voice quality is another, widely studied, communication construct that has been proposed to serve as a cue for gender attribution. There are multiple ways to perceptually characterize voice quality; however, breathiness and roughness seem to be of particular interest in the context of gender attribution. Feminine-sounding voices have been characterized by increased breathiness,[33,44−48] whereas rougher or "croakier" voices (ie, self-rated vocal fry) have been associated with masculine-sounding voices.[10]

Several measures have been used in previous research to quantify these subjective constructs. For example, measures of cycle-to-cycle variability in vocal fold timing (ie, jitter, relative average perturbation) and amplitude (ie, shimmer) have been related to perceived vocal harshness or hoarseness,[33,49,50] noise-harmonic ratio (NHR) with roughness,[51] and soft phonation index (SPI) with breathiness.[51]

There are limited studies investigating these acoustic parameters as a function of gender attribution amongst transgender women. Of note, shimmer previously has been moderately negatively correlated (P = 0.076) with listener femininity ratings[18] and SPI has been shown to contribute to an individual being attributed as a woman.[52] In other cases, vocal perturbation measures (ie, jitter, shimmer, NHR) either were excluded from the study or found to correlate weakly with femininity ratings due to a lack of variability in these acoustic measures.[18,27,33]

In addition to its potential relationship with femininity ratings, shimmer also has been negatively related to naturalness ratings in the pretraining voices of transgender women (n = 25).[27] Although not found to be a significant predictor of naturalness ratings on its own, shimmer did significantly predict these ratings when included in a model with F2 of /a/ and minimum frequency in a maximum phonational frequency range task.

Overall, it appears that voice quality may contribute to perception of gender, femininity−masculinity, and/or vocal naturalness, but additional research is needed to more clearly describe this contribution. Given previous research results and auditory-perceptual characterizations of masculine and feminine voices, shimmer, NHR, and SPI may be the acoustic measures most worthwhile to investigate.

In summary, the existing literature provides compelling evidence for the role of f_o and vocal tract resonance in gender attribution; however, it also is clear that other aspects of the voice serve as cues for gender and masculinity−femininity to others. Research has been somewhat equivocal with respect to the role and relative strength of other acoustic variables as predictors of these auditory-perceptual constructs.

Studies have looked at certain of these acoustic variables either in isolation or in groupings related to a few areas of voice (e.g., f_o, intonation, resonance). A more comprehensive investigation into the relative contribution of acoustic variables—representing a number of different aspects of voice and speech—to gender attribution is needed. Furthermore, although it is becoming increasingly common, not all studies have included perceptual ratings. When included, these ratings often have focused on either gender attribution or perceived masculinity−femininity but not always both. Information also is lacking describing how acoustic variables contribute to a natural-sounding voice: only one study has included this auditory-perceptual measure with transgender communicators.[27]

Purpose

The purpose of this study is to identify the most salient set of acoustic predictors of gender attribution, perceived femininity, and perceived vocal naturalness in a conversational context. Findings will help inform voice and communication feminization training so that goals are centered on the strongest cues to gender in conjunction with the areas identified as most important to the individual client. Improving our understanding of the predictive relationships between these acoustic variables and perception of femininity and vocal naturalness will guide further refinement of gender presentation so that speakers are attributed not only as their affirmed gender but also in a way that is congruent with their specific identity (e.g., a very feminine woman, somewhat masculine woman) without sounding unnatural. In this way, clinicians will be better able to tailor training so that it is appropriate to the self-identified goals of the client.

Specific questions related to these research purposes are: What acoustic measures predict: (1) gender attribution (as man, woman, or ambiguous in gender)?; (2) perceived masculinity−femininity?; and (3) perceived vocal naturalness?

METHODS

Participants

Two groups of participants were recruited to take part in the study: communicators and raters. This study received ethical approval from the University of Alberta Health Research Ethics Board and operational approvals from the Northern Alberta Clinical Trials and Research Centre and the Covenant Health Research Centre prior to participant recruitment.

Communicators

A total of 42 communicators were recruited for the study: 22 transgender women as well as 10 (each) cisgender men and women. Inclusion criteria required that all communicators be fluent English speakers and not have a neurogenic communication disorder. Additional inclusion criteria for transgender women communicators required that they identify as transsexual or transgender and be living in their affirmed gender role (ie, woman) the majority of the time (ie, at least 80% of the time) for at least 6 months. One transgender participant was excluded from the study due to the presence of a non-North American accent: the authors believed the accent would impact subsequent acoustic measures and may have acted as a covariate during the rating. Another transgender participant was excluded as a result of failing to meet the inclusion criterion of living in the affirmed gender role the majority of the time. These exclusions resulted in a final sample size of n = 40.

The sample size (for transgender women) used in the present study is the same as that used by Wolfe, Ratusnik, Smith, and Northrop[24] and is representative of the number of transgender women that could reasonably be expected to be recruited from the Edmonton area. Cisgender men and women were included to ensure adequate variability in the sample and each of these participants was age-matched with one of the transgender participants. All but two age matches were within 2 years and the remaining two matches were within 4 years. There were no significant differences in age between gender identity groups as determined by a one-way between-subjects Analysis of Variance $F(2,37) = 0.010$, $P = 0.990$. Demographic information for all communicators can be found in Table 1. Transition-related information for transgender communicators is summarized in Table 2.

Information about smoking history was collected due the potential of smoking to affect voice quality.[54] History of smoking was similar across gender groups and all groups had members who had succeeded in smoking cessation by the time of data collection. Of note, no cisgender women continued to smoke whereas there were similar, small numbers of active smokers in the other two groups. Overall, relatively few of the communicators were smokers.

The transsexual voice questionnaire (TVQ)^{MtF} is a self-report tool developed specifically for transgender women as a means of measuring the impact of their voices on their day-to day-lives.[55] The average total score for this group of transgender women fell 15 points below the midpoint, indicating that, in general, these participants had some negative experiences related to voice function and/or social participation but these difficulties were not necessarily happening on a very frequent basis.

Cisgender communicators were recruited using poster advertisements and word-of-mouth. Transgender communicators were recruited from gender and voice clinics, support groups, gender and sexual minority organizations, and the general public using poster advertisements, word-of-mouth, and a media release. Letters of invitation to participate also were sent to individuals who had been referred for voice and communication feminization services through the voice program at the local rehabilitation hospital. Individuals who were interested in participating and who were awaiting or were actively receiving services were asked to contact another author in order to prevent feelings of coercion as the first author was the clinician responsible for providing said services. Individuals who already had received services or who were discharged without being seen were asked to contact the first author if they were interested in participating.

Raters

A total of 20 raters were recruited for participation in the study from the University of Alberta and Edmonton area through convenience sampling and word-of-mouth and represented the naïve communication partners encountered in the communicators' daily lives. Half of the raters identified as women and half identified as men. The mean age (n = 19) was 29.26 and ranged from 18 to 46 years. One rater declined to provide her age. Raters were required to (a) be proficient in English; (b) pass a pure-tone hearing screening at 25 dB HL for 500, 1000, 2000, and 4000 Hz bilaterally; (c) have no uncorrected concerns with vision per self-report; (d) have no specific training in listening; (e) have no more than incidental experience listening to or communicating with persons having communication disorders; and (f) have no identified language, learning, or cognitive disabilities per self-report.

Most of the raters had at least some undergraduate education and all had at least a high school diploma. The level of education attained (including partial completion) ranged from community college to advanced graduate level. Raters varied with respect to the amount of their family, friends and/or close colleagues who identified with the greater lesbian, gay, bisexual, transgender, and queer (LGBTQ) community; however, the majority had at least a few. Raters were asked to indicate their sexual orientation on a nine-

TABLE 1.
Demographic Information of Communicators

Demographic Variable	Transgender Women (n = 20)	Cisgender Women (n = 10)	Cisgender Men (n = 10)	All Communicators (n = 40)
Mean age in years (SD)	41.20 (14.38)	40.90 (14.76)	40.40 (15.93)	40.93 (14.48)
Age range in years	16–69	24–68	18–70	16–70
History of smoking (%)	45	40	30	40
Current smoker (%)	15	0	20	12.5

TABLE 2.
Transition-Related Information for Transgender Communicators

Transition-Related Characteristic	N	Mean (SD)	Min−Max	n
Time living in feminine gender role at least part-time (years)	20	9.08 (11.71)	1.0−52.0	-
Time living "full-time" in feminine gender role (years)	20	7.63 (11.72)	0.75−52.0	-
Length of SLP services (months)	19	1.46 (3.42)	0.0−12.0	-
Time since SLP services (months)	8	45.00 (54.81)	0.0−156.0	-
Receiving HRT	20	-	-	18
Underwent GAS	20	-	-	8
TVQ^{MtF} Total Score	20	60.70 (18.96)	31−91	-

Note. One participant received SLP services but could not recall for how long. Twelve participants had not received SLP services at the time of participation. GAS, gender affirmation surgery; HRT, hormone replacement therapy; TVQ^{MtF}, transsexual voice questionnaire (male-to-female).[53] Minimum and maximum possible for total score = 30−120.

point scale anchored with 1 = very heterosexual and 9 = very homosexual as in Hancock and Pool.[56] The median sexual orientation for the whole group was straight according to the criteria used by those authors (ie, less than three was considered straight); however, half of the raters scored themselves as a three or greater, which would be considered nonstraight by the same criteria. Raters were asked to provide information about their exposure to the LGBTQ community and their sexual orientation because femininity ratings made of transfeminine speakers have been related to sexual orientation[56] and because exposure to sexual and gender diverse communities may result in broader conceptualizations of gender expression.

Data collection for communication variables

Recording equipment and set-up

Data collection took place at the Syncrude Centre for Motion and Balance at the Glenrose Rehabilitation Hospital in Edmonton. Upon arrival, participants provided consent to participate in the study and subsequently were fitted with recording equipment. A Shure Mx185 condenser microphone was attached to the communicator's forehead at a fixed distance of 10 cm from the mouth and connected to an audio-buddy amplifier and 32-bit HP laptop computer. Monaural audio recordings were collected using TF32 software[57] at a sampling rate of 44.1 KHz. Gain was calibrated on the audio-buddy amplifier prior to recording the research tasks. No further adjustments were made once those tasks commenced. Communicators were seated on a raised stool and were positioned facing a listener (the first author) as if in a conversational exchange.

Speaking tasks

Audio recordings were taken of the communicators while they retold the story of the Pink Panther cartoon, "In the Pink of the Night."[58] A story retell task was chosen to control for the use of potentially-confounding vocabulary (e.g., my wife) that may inadvertently be used in a traditional narrative task while still allowing for more conversational speech patterns than may be elicited during reading.

Two additional speaking tasks were performed and recorded for use in obtaining specific acoustic measures. First, communicators were asked to read the carrier phrase, "Say hVd again" (with the vowels /i/, /ɑ/, and /u/), as in Gelfer and Bennett.[13] They read the carrier phrases five times in total. This task was chosen to control phonetic context during vowel production in connected speech. Second, communicators were recorded while sustaining the vowel /ɑ/ (as in "pop") for approximately 5 seconds over five trials. This task was included for use in obtaining vocal stability measures, which provide information about nonvolitional cycle-to-cycle variability in frequency and SPL.[59] As such, sustained vowels provide an ideal context for capturing vocal stability measures because they do not include volitional variations in frequency and SPL found in connected speech secondary to intonation.[59]

Calibration tones were recorded at the end of each data collection session, as in Fox and Boliek,[60] for later use in calculating SPL. Calibration tones were generated at 440.0 Hz using a tone generator application on a smartphone. The speaker volume on the smartphone was set to maximum and the speaker was placed at the center of the communicator's mouth (ie, the same point from which microphone distance was measured).

Editing cartoon descriptions

Communicators were variable in the length and completeness of their cartoon descriptions. Consequently, the narratives were edited to control for length and linguistic complexity. Linguistic complexity was measured in t-units where a t-unit is defined as "one main clause plus whatever subordinate clauses happen to be attached to or embedded within it"[61] (p.305). Each sample was edited to be between 30 and 45 seconds in length and contain 8−12 t-units. Care was taken to exclude vegetative acts (e.g., laughing) from the sample. These edited descriptions were used for the rating phase of data collection and for the majority of the acoustic measures.

Acoustic measures

A total of 13 acoustic measures initially were captured: mean f_o (Hz); mean minimum frequency (Hz); mean maximum frequency (Hz); mean RoS (syllables/s); mean SPL (dBSPL); mean ST range, percent upward intonation shifts; mean first, second, and third vowel formants for the vowel /i/ (Hz); mean NHR; mean shimmer (%); and mean SPI. These served as the predictor (ie, independent) variables in the statistical analyses. Customized PRAAT[62] scripts[63,64] were used to obtain all the acoustic measures with the exception of ST range and SPI, which were obtained using the Real-Time Pitch and Multi-Dimensional Voice Program modules in Multi-Speech,[65] respectively. The first seven measures were taken from the edited cartoon description, averaged across t-units. Vowel formant frequencies were measured in the context of the carrier phrase at the midpoint of the vowel (averaged across the five productions), and the three vocal stability measures were obtained from the middle one second of the sustained /ɑ/ (averaged across the five productions).

Intonation shifts were defined[32] as "a change in frequency, with or without interruption of phonation, of at least two semitones" (p.26). For each t-unit, the script calculated the difference between the maximum and minimum frequency and converted the value to semitones using the formula: $30.86314 * \log10(f_o max/f_o min)$ as in Hancock, Colton, and Douglas.[17] If this value was at least two ST, it was counted as an intonation shift. The direction of the shift was determined by calculating the difference between the time of the maximum and minimum f_o values. If the value was positive, it was counted as an upward shift and if negative, it was counted as a downward shift. Percent upward intonation shifts subsequently were calculated by dividing the number of upward shifts by the total number of shifts and multiplying by 100.

All acoustic measures were repeated for eight (20%) randomly selected participants for reliability purposes. Intrameasurer reliability was determined using repeat measures made by the first author whereas intermeasurer reliability was calculated using repeat measures made by the first author and two masters-level SLP students.

Data collection for auditory-perceptual ratings

Rating procedure and rating scales

Data collection took place in a quiet room either at the Clinical Sciences Building at the University of Alberta or at the Glenrose Rehabilitation Hospital. After providing consent and demographic information, the raters were provided with a standard training protocol via a slideshow presentation. The training introduced raters to relevant concepts (eg, masculinity−femininity, vocal naturalness) and the rating scales, and provided them with examples of different voices (eg, voice attributed as man/woman/ambiguous in gender).

Data collection was accomplished using a customized software program, "Gender Finder"[66] that randomly presented the edited cartoon description samples from all 40 communicators followed by 10% of the files repeated for reliability purposes. Raters were asked to make three ratings for each audio file: gender, perceived masculinity −femininity of the communicator, and perceived vocal naturalness. Gender attribution ratings were made by selecting one of three options from a drop-down menu: man, woman, or can't decide. The third option "can't decide" was included to expand the binary choice typically provided when inquiring about gender and to capture those presentations that are attributed gender-neutrally since that is a desired outcome for some transgender people. These options were believed to adequately represent gender attributions made by the general public. Masculinity−femininity ratings provided additional information about the spectrum of gender presentations and how they are perceived. Masculinity−femininity and naturalness ratings both were made using direct magnitude estimation scales without modulus (DME-WM) where higher numbers represented more feminine and more natural voices and lower numbers represented more masculine and more unnatural voices, respectively. This type of scale requires raters to assign a number to the first stimulus item to which they are presented and rate subsequent stimuli in comparison to the first. For example, if the first item was assigned 200 and the second item was perceived to be twice as feminine, it would be assigned 400. Conversely, if the second file was perceived to be half as feminine, it would be assigned 100. The naturalness scale was structured in an analogous fashion. Raters were permitted to use any number of their choosing as long as it was not a negative number or zero. Ratings were entered directly into Gender Finder via a rating window that appeared on the screen following each audio file.

DME scales do not have fixed end points and are considered to provide ratio-level data.[67,68] Mathematical conversions were performed (ie, modulus equalization) with the DME-WM ratings prior to their use in statistical analyses to remove the variability secondary to the use of different moduli between raters.[69] That is, ratings were mathematically converted to be on the same scale with the same modulus. Modulus equalization also creates a normal distribution, allowing for the use of parametric statistics.

Raters were given an opportunity to practice using the DME-WM rating scale during the training exercise by assigning numbers to lines of various lengths similar to Snow and Williges.[69] They were provided with additional experience prior to commencing the research trials during a practice session with the Gender Finder program. The software was written to have a practice module as well as an experimental module. The practice module was the same as the experimental module except that it used a smaller set of unique stimuli and rating data were not stored. The recordings used for the practice module were of two pilot participants from the research team and the two communicators who did not meet inclusion criteria. The practice module could be repeated as many times as the rater desired until they were comfortable with the program and the tasks. No rater completed the practice module more than once.

Raters were seated at a desk or table directly in front of a laptop computer that was loaded with the Gender Finder software and provided with a set of Bose QuietComfort 35 noise-cancelling headphones. All training, practice, and research trials were completed on the laptop computer. Raters were instructed to adjust the volume on the laptop to a comfortable level during the practice trials and were not allowed to make further adjustments once the research trials commenced. They also were provided opportunities to ask questions after the training and practice activities. Research trials began following the practice module. The auditory-perceptual ratings took approximately 35−50 minutes to complete.

Once all ratings were completed, raters were asked to answer a two-item post-rating questionnaire about their exposure to the LGBTQ community and their sexual orientation. They then were debriefed about the study and the reason for asking the postrating questions. Raters were not advised of the inclusion of transgender women communicators prior to their participation in the study.

Auditory-perceptual ratings

Composite gender attribution, masculinity−femininity, and vocal naturalness ratings were calculated for each communicator. Gender attribution was assigned according to the criteria used in Gelfer and Bennett[13] where an individual was considered to be consistently attributed as a particular gender if they were attributed that gender at least 80% of the time. If none of the three gender choices was selected at least 80% of the time, that communicator was assigned "can't decide" as their gender was ambiguous to the raters. Masculinity−femininity and vocal naturalness ratings were averaged across raters to arrive at a mean overall rating. These ratings served as the criterion (ie, dependent) variables in the statistical analyses.

Statistical analyses

The first research question regarding acoustic predictors of gender attribution was answered using logistic regression due to the categorical nature of the criterion variable. Standard multiple linear regressions were used to identify sets of acoustic predictors of perceived masculinity−femininity and vocal naturalness ratings and, thus, answer the second and third research questions. Multiple regressions were used because masculinity−femininity and naturalness ratings were continuous variables providing ratio-level data.

Given the large number of predictors and the relatively small sample size, it was necessary to decrease the number of predictors used for each regression analysis to maximize the participant: predictor ratio. This culling of predictors was accomplished by identifying and removing highly correlated predictors and by using a purposeful selection of covariates model-building strategy as recommended by a consulting statistician and by Hosmer, Lemeshow, and Sturdivant.[70] Purposeful selection identifies the most parsimonious set of predictors for use in further analyses. Variables that were highly correlated with other variables were removed from the potential predictor set for all of the analyses. Purposeful selection of covariates was performed separately for each of the regression analyses.

RESULTS

Reliability of acoustic measures

Intra- and intermeasurer reliability were examined for all 13 acoustic measures. Intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals (CI) were calculated using SPSS statistical software version 24 based on a two-way mixed effects model with absolute agreement. Results are presented in Table 3. Based on the 95% CI and using criteria suggested by Koo and Li,[71] intrameasurer reliability was excellent for all measures except ST range and percent upward intonation shifts. Reliability ranged from

TABLE 3.
Intraclass Correlation Coefficients (ICC) and 95% Confidence Intervals (CI) for Acoustic Measures

	Intrameasurer Reliability		Intermeasurer Reliability	
Acoustic Measure	ICC	95% CI	ICC	95% CI
f_o	1.0	[0.999, 1.0]	0.998	[0.995, 1.0]
Minimum frequency	0.999	[0.994, 1.0]	0.963	[0.887, 0.992]
Maximum frequency	1.0	[0.998, 1.0]	0.998	[0.994, 1.0]
F1 /i/	1.0	[0.998, 1.0]	0.949	[0.844, 0.989]
F2 /i/	0.983	[0.925, 0.997]	0.929	[0.792, 0.984]
F3 /i/	0.998	[0.987, 1.0]	0.989	[0.964, 0.998]
ST range	0.978	[0.891, 0.996]	0.91	[0.741, 0.979]
% Upward intonation shifts	0.699	[(−)0.562, 0.940]	-	-
SPL	0.999	[0.994, 1.0]	0.858	[0.617, 0.966]
Shimmer	0.996	[0.979, 0.999]	0.984	[0.947, 0.996]
NHR	0.999	[0.993, 1.0]	0.979	[0.935, 0.995]
SPI	0.997	[0.984, 0.999]	0.985	[0.953, 0.997]
RoS	0.995	[0.975, 0.999]	0.997	[0.992, 0.999]

good to excellent for ST range and percent upward intonations shifts were not reliable. The reason for lack of reliability was thought to be due to the method of measurement. That variable was excluded from subsequent analyses, including intermeasurer reliability.

Intermeasurer reliability was excellent for seven of the remaining 12 measures. Mean minimum frequency and F1 /i/ ranged from good to excellent reliability whereas F2 /i/, ST range, and mean SPL ranged from moderate to excellent reliability.

Reliability of auditory-perceptual ratings

Intrarater reliability was calculated between the first and second ratings of the four repeated files using percent agreement for gender attribution and two-way mixed effects model ICC estimates with absolute agreement for masculinity−femininity and vocal naturalness ratings. Results indicated that intrarater reliability was excellent for gender attribution (% agreement = 92.5%), excellent for masculinity−femininity ratings (ICC = 0.946, 95% CI = 0.916−0.965), and ranged from good to excellent for naturalness ratings (ICC = 0.875, 95% CI = 0.804−0.921).

Inter-rater reliability was calculated between the ratings made by each of the 20 raters using Fleiss' Kappa for gender attribution and two-way mixed effects model ICC estimates with absolute agreement for masculinity−femininity and vocal naturalness ratings. Results indicated that inter-rater reliability was good for gender attribution ($\kappa = 0.81$), excellent for masculinity−femininity ratings (ICC = 0.94, 95% CI = 0.909−0.964) and ranged from good to excellent for naturalness ratings (ICC = 0.854, 95% CI = 0.777−0.913).

Correlations between acoustic variables

Bivariate Pearson correlation coefficients were calculated between each of the 12 remaining acoustic variables. Statistically significant correlations ($r = 0.887{-}0.969$, $P < 0.01$) were revealed between f_o, mean minimum frequency, and mean maximum frequency. Given the wealth of literature supporting f_o as a cue to gender attribution, it was selected to remain in the set and the other two variables were excluded.

No other correlation coefficients exceeded 0.8; however, significant correlations were identified between F1 and F2 of /i/ ($r = 0.339$, $P < 0.05$) and F2 and F3 of /i/ ($r = 0.713$, $P < 0.01$). In an effort to further reduce the number of potential predictors, an average was taken of the three vowel formants of /i/, similar to Gelfer and Bennett,[13] resulting in the creation of a single measure to represent vocal tract resonance (ie, "average formant frequency"). Descriptive statistics for each of the remaining eight acoustic measures along with the DME-WM scale ratings are presented in Table 4. Of note, SPI data were missing for five of the communicators (12.5%) because all five trials of the sustained vowel yielded an error response when analyzed in Multi-Speech.[65] In addition, one rater (R12) did not rate

TABLE 4.
Descriptive Statistics for Acoustic Measures, Femininity and Naturalness Ratings

Acoustic Measure	All Communicators (n = 40)		Attributed Man (n = 25)		Attributed Woman (n = 11)		Attributed Ambiguously (n = 4)	
	Mean (SD)	Min−Max	Mean (SD)	Min−Max	Mean (SD)	Min−Max	Mean (SD)	Min−Max
f_o (Hz)	147.73 (40.04)	96.22−230.58	122.93 (21.67)	96.22−196.28	194.60 (22.14)	150.46−230.58	173.84 (36.84)	146.06−227.93
Average formant frequency (Hz)	1952.22 (141.04)	1692.99−2344.64	1887.16 (95.55)	1692.99−2116.39	2101.73 (138.82)	1875.80−2344.64	1947.66 (63.83)	1864.57−2000.87
Shimmer (%)	0.035 (0.017)	0.014−0.077	0.041 (0.018)	0.014−0.077	0.037 (0.012)	0.014−0.058	0.024 (0.004)	0.018−0.028
NHR	0.017 (0.12)	0.002−0.137	0.021 (0.03)	0.002−0.137	0.012 (0.01)	0.003−0.034	0.007 (0.01)	0.002−0.011
SPI[a]	18.39 (8.84)	4.73−34.36	16.95 (7.80)	4.73−34.06	19.13 (11.54)	7.09−34.36	24.64 (6.10)	16.20−29.88
SPL (dBSPL)	74.68 (3.50)	69.39−82.43	75.02 (3.57)	69.76−82.43	73.77 (3.86)	69.39−82.19	75.13 (1.88)	73.46−76.85
RoS (syllables/s)	3.70 (0.56)	2.36−4.81	3.69 (0.52)	2.60−4.81	3.94 (0.56)	2.81−4.53	3.15 (0.56)	2.46−3.68
ST range	9.0 (2.59)	4.58−16.63	9.21 (2.65)	5.17−16.63	8.23 (2.71)	4.58−14.25	9.79 (1.72)	7.63−11.67
Masculinity−femininity rating	21.51 (7.48)	10.72−35.71	-	-	-	-	-	-
Vocal naturalness rating	17.64 (1.98)	11.12−19.94	-	-	-	-	-	-

[a] SPI had missing values and resultant sample sizes were: all communicators = 35; attributed man = 22; attributed woman = 9; attributed ambiguously = 4.

naturalness for any communicator. Two additional naturalness ratings (from R2 and R7) were missing and likely were due to mouse click errors during rating. These omissions represented 5.25% of the total naturalness rating data.

Correlations between transition-related factors and auditory-perceptual ratings

Bivariate Pearson correlation coefficients were calculated between a number of personal and transition-related factors and masculinity−femininity ratings for the transgender communicators. There were no significant relationships between the ratings and time since transition commenced ($r = 0.225$, $P = 0.341$, two-tailed), length of time living full-time in the feminine gender role ($r = 0.212$, $P = 0.370$, two-tailed), length of voice and communication training ($r = -0.207$, $P = 0.394$, two-tailed), time since the completion of training ($r = 0.674$, $P = 0.067$, two-tailed), or TVQ^{MtF} total score ($r = -0.381$, $P = 0.097$, two-tailed).

Acoustic predictors of gender attribution

A multinomial logistic regression was used to model the relationship between a set of acoustic predictors and gender attribution (man, woman, can't decide) (n = 40). Purposeful selection univariable analyses (ie, 1-way Analysis of Variance or Kruskal-Wallis test) identified four significant variables ($\alpha = 0.25$): f_o, average formant frequency, shimmer, and RoS. These variables were retained for inclusion in the regression analysis, resulting in a sample size of n = 10 per variable.

A total of 25 (62.5%) communicators were attributed as men, 11 (27.5%) were attributed as women and raters could not decide the gender of four (10%) of the communicators; therefore, the referent group for the analysis was "attributed as a man" because it was the most frequently attributed gender. The Goodness of Fit test was not significant χ^2 (70, N = 40) = 47.437, $P = 0.995$, indicating that the model fit the data well. The final model (ie, the model including the predictors) predicted gender attribution significantly better than the model that contained only the intercept χ^2 (8, N = 40) = 47.437, $P < 0.001$. Pseudo R^2 (Nagelkerke) revealed that 83.9% of the variance in gender attribution was explained by the model. As is shown in Table 5, f_o and RoS were statistically significant; however, only f_o had a significant parameter for comparing attributed man and attributed woman (Wald $\chi^2 = 3.951$, df = 1, $P = 0.047$). The odds that an individual would be attributed as a woman rather than a man increased by a factor of 1.133 (95% CI = 1.002−1.281) for every unit increase in mean f_o. Overall the model correctly predicted gender attribution for 90.0% of cases.

The analysis subsequently was repeated including only f_o as a predictor. Similar results were obtained except that in this case, the model explained 71.2% of the variance in gender attribution and significant parameters were revealed for comparisons between attributed man and both attributed woman (Wald $\chi^2 = 9.236$, df = 1, $P = 0.002$) and can't decide

TABLE 5.
Unique Contribution of Predictors in the Multinomial Logistic Regression (n = 40)

	Likelihood Ratio Test	
Predictor	χ^2 (df = 2)	P ($\alpha = 0.05$)
F_o	15.80	< 0.01*
Average formant frequency	1.59	0.451
Shimmer	3.67	0.159
RoS	8.42	0.015*

Note. χ^2 = the increase in −2 log likelihood when the predictor is removed from the full model. Statistically significant results are marked with *.

(Wald $\chi^2 = 5.792$, df = 1, $P = 0.016$). For every unit increase in mean f_o, the odds that an individual would be attributed as a woman or ambiguous in gender increased by a factor of 1.107 (95% CI = 1.037−1.182) and 1.079 (95% CI = 1.014−1.149), respectively. Mean f_o was 122.93 Hz (SD = 21.67) for communicators who were attributed as men, 194.60 Hz (SD = 22.14) for those attributed as women, and 173.84 Hz (SD = 36.84) for those whose gender was ambiguous to raters (Figure 1). This model correctly predicted gender attribution in 85% of cases.

Acoustic predictors of masculinity−femininity ratings

Multiple linear regression was used to identify a set of acoustic predictors of masculinity−femininity ratings. Five predictors were found to be significant through purposeful selection procedures (ie, simple regression) and thus were retained for the multiple regression analysis using the standard method. These variables included f_o, average formant frequency, shimmer, NHR, and SPL.

The results revealed a significant model: $F(5,34) = 30.154$, $P < 0.001$ that explained 78.9% of the variance in masculinity−femininity ratings (adjusted $R^2 = 0.789$). Regression

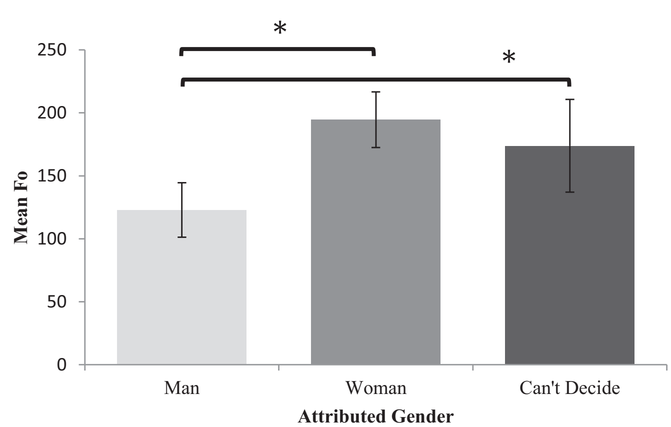

FIGURE 1. Mean f_o for each attributed gender group. Error bars represent +/− 1 SD. Statistically significant group comparisons are marked with *.

coefficients for the predictor variables are summarized in Table 6. F_o, average formant frequency, and SPL were significant predictors of masculinity–femininity ratings. F_o and average formant frequency were positively related to the ratings whereas SPL was negatively related: Higher speaking frequencies and vocal tract resonance characteristics were associated with more feminine/less masculine ratings and voices with greater SPL were associated with less feminine/more masculine ratings.

The multiple regression was repeated using only the three significant predictors in order to obtain new regression coefficients. These also are presented in Table 6. Differences in beta coefficients did not exceed 20%; therefore, there likely was not an interaction with a variable that was excluded.[70] These coefficients revealed that f_o made the greatest contribution to the model, followed by average formant frequency, and finally, SPL. For every change of one SD in mean f_o (ie, 40.04 Hz), we would predict a change of 0.557 SD in masculinity–femininity ratings or about four points. Similarly, for every one SD change in average formant frequency and SPL, we would expect a change of 0.395 SD (ie, approximately three points) and −0.252 SD (ie, approximately 2 points) in masculinity–femininity ratings, respectively.

Acoustic predictors of vocal naturalness ratings

The same statistical procedures were used to identify the set of acoustic predictors of vocal naturalness as were used for identifying predictors of masculinity–femininity ratings. Four predictors were selected to enter into the multiple regression model: f_o, average formant frequency, NHR, and RoS. The results of the multiple linear regression revealed a significant model: $F(4,35) = 6.579$, $P < 0.001$ that explained 36.4% of the variance in vocal naturalness ratings (adjusted $R^2 = 0.364$). Significant predictors included f_o, average formant frequency, and RoS. f_o was negatively related to vocal naturalness. The other two predictors were positively related. Regression coefficients are presented in Table 7.

TABLE 6.
Unstandardized and Standardized Regression Coefficients for the Predictors Entered into the Model

Predictor	B	SE B	β	P (α = 0.05)
F_o	0.101[a]	0.019[a]	0.539[a]	< 0.001*,[a]
	0.104[b]	0.019[b]	0.557[b]	< 0.001*,[b]
Average formant frequency	0.019[a]	0.005[a]	0.366[a]	0.001*,[a]
	0.021[b]	0.005[b]	0.395[b]	< 0.001*,[b]
Shimmer	−83.175[a,b]	50.692[a,b]	−0.191[a,b]	0.110[a,b]
NHR	48.563[a,b]	33.208[a,b]	0.151[a,b]	0.153[a,b]
SPL	−0.652[a]	0.172[a]	−0.305[a]	0.001*,[a]
	−0.538	0.160[b]	−0.252[b]	0.002*,[b]

[a] Coefficients obtained from the first model including all five predictor variables.
[b] Coefficients obtained from the second model including only the three predictors that were significant in the first model.
Note. Significant predictors of masculinity–femininity ratings are marked with *.

TABLE 7.
Unstandardized and Standardized Regression Coefficients for the Predictors Entered into the Model

Predictor	B	SE B	β	P (α = 0.05)
F_o	−0.029[a]	0.009[a]	−0.592[a]	0.002*,[a]
	−0.032[b]	0.009[b]	−0.650[b]	0.001*,[b]
Average formant frequency	0.009[a]	0.002[a]	0.641[a]	0.001*,[a]
	0.009[b]	0.002[b]	0.642[b]	0.001*,[b]
NHR	15.462[a,b]	11.511[a,b]	0.182[a,b]	0.188[a,b]
RoS	1.091[a]	0.457[a]	0.312[a]	0.022*,[a]
	1.035[b]	0.460[b]	0.295[b]	0.031*,[b]

[a] Coefficients obtained from the first model including all four predictor variables.
[b] Coefficients obtained from the second model including only the three predictors that were significant in the first model.
Note. Significant predictors of vocal naturalness ratings are marked with *.

Once again, the multiple regression was repeated with only the three significant predictors entered into the model. The revised regression coefficients also are summarized in Table 7. Interactions with NHR are unlikely given the small differences in beta coefficients.

DISCUSSION

This study aims to identify the most salient set of acoustic predictors of gender attribution as well as perceived masculinity–femininity and vocal naturalness for the purpose of contributing to the evidence base informing voice and communication modification training programs for transgender individuals. Each of these aims was addressed with a separate research question. The discussion will focus primarily on significant findings.

Acoustic predictors of gender attribution

The results of the multinomial logistic regression revealed a significant predictive model and identified f_o as the sole acoustic predictor that significantly changed the odds of being attributed as a woman or ambiguous in gender rather than as a man. As mean f_o increased, so too did the chances of being attributed a gender other than man.

Mean f_o for participants attributed as men and women was consistent with reported norms for cisgender speakers.[59] Mean f_o for the group attributed as gender ambiguous was within the suggested gender ambiguous or gender neutral range.[28] Examination of the minimum and maximum values (Table 4) revealed that mean f_o for participants attributed as women did not fall below 150 Hz. This value is at the lower boundary of the gender-neutral range and is slightly lower than minimum threshold values previously suggested for a transgender woman to be attributed as a woman, approximately 155–160 Hz.[24,72] The minimum value for mean f_o for the ambiguously attributed group

(146.06 Hz) was just below the gender neutral frequency range. As such, targeting a mean f_o that falls within the cisgender woman or gender neutral range appears to be beneficial for minimizing attributions as a man based on the voice.

Maximum mean f_o values also were informative. That of participants attributed as men was approximately 30−35 Hz lower than those of the other two attributed gender groups; however, it fell within the cisgender woman range. This result provides further evidence for the assertion that modifying f_o alone or speaking at a mean f_o within the cisgender woman range may not be sufficient for consistent attribution as a woman. In this study, 71.2% of the variance in gender attribution could be accounted for by f_o, leaving 28.8%, almost a third of the variance, remaining to be explained by other variables.

It is possible that other acoustic variables may have emerged as significant predictors had the sample had more variability with more even distribution across the three attributed gender groups. Almost 2/3 of the communicators who participated in this study were attributed as men, including 16 of the 20 transgender communicators (one was attributed as a woman and three were attributed as gender ambiguous). Only about 1/4 were attributed as women, and 1/10 were attributed ambiguously. Two recent review articles (one with meta-analysis) provide direction regarding which of the acoustic variables included in the present study may have emerged as significant predictors of gender attribution. These variables include vowel formant frequencies, f_o variability, minimum and maximum f_o, intonation measures such as directional shifts and ST range, SPL, and acoustic correlates of voice quality (eg, SPI).[15,11] Each of these variables has evidence to support its status as a voice marker of speaker gender.

Methodological factors also may have limited the identification of significant predictors in addition to the lack of sufficient variability in communicator voice and speech characteristics. Several of the listed variables were measured in the present study but not included in the regression analyses either because they were measured for descriptive purposes only (ie, f_o variability), the measurements were not reliable (ie, percent upward intonation shifts), or they were excluded due to multicollinearity with other variables (ie, minimum and maximum f_o). It would be valuable to systematically explore the salience of these predictors along with f_o amongst gender diverse speakers.

Despite the limitations in communicator variability and acoustic variable inclusion, the results of the present study nevertheless are consistent with previous research in demonstrating that f_o is a strong cue to gender attribution[11] and support the common and recommended practice of targeting raising f_o in voice and communication feminization training.[73,74]

Acoustic predictors of masculinity−femininity ratings

The results of the multiple linear regression analyses identified f_o, average formant frequency, and SPL as significant predictors of masculinity−femininity ratings. Higher values of f_o and average formant frequency were associated with more feminine ratings, whereas higher SPL values were related to more masculine ratings. Together, these variables accounted for most of the variability in masculinity−femininity ratings (adjusted $R^2 = 0.789$) with f_o contributing the most to the model, followed by average formant frequency, and finally SPL. These results confirm and extend previous research and provide novel findings.

First, the identification of f_o as a predictor of masculinity−femininity ratings is not surprising given the results presented in the previous section (4.1) and in the existing literature. The results of this study confirm f_o as an important cue to perceived masculinity−femininity of the voice and further support the goal of raising f_o in training when aiming to achieve a more feminine-sounding voice.

Second, the results also are consistent with previously published research that reported vocal tract resonance to be a salient cue for gender attribution in the context of connected speech[13,14] and extends those findings by revealing the same relationship with ratings of masculinity−femininity and including transgender communicators. In addition, the results of the present study are similar to those reported in an unpublished master's thesis.[75] Both studies found positive relationships between vowel formant frequencies and masculinity−femininity ratings. Dahl found stronger relationships for cisgender participants than for transgender ones; however, small sample size (n = 12 transgender communicators, n = 10 cisgender communicators) and an ordinal rating scale may have limited the results.

It is interesting and important to note that formant frequency emerged as a significant predictor in this study even though it was averaged across the first three formants of /i/. This result suggests that vocal tract resonance, overall, is salient for conveying one's masculinity or femininity to others. Average formant frequency has been linked to gender attribution in the past and was suggested to better represent the perceptual experience of listeners than individual vowel formants.[13,76] The previous studies averaged the first three formants for two or three vowels (ie, /i/, /ɑ/, /ɛ/, and/or /u/) whereas the present study used only /i/. Since the results were the same when the composite measure was obtained using only one vowel as when two or three vowels were used, perhaps clinicians could use the single vowel composite for assessment and outcome measurement in order to save time.

Taken together, these results suggest that training communication behaviors that decrease the length of the vocal tract and/or size of the resonating chambers (eg, lips spreading, anterior tongue carriage) and thus, raise the frequency of the vowel formants overall, may result in meaningful training outcomes. Such outcomes were observed in an intervention study that targeted these behaviors[25] as well as in two other studies investigating postintervention changes following a course of voice feminization training that included facilitation of oral resonance patterns.[20,77] All three studies found that participants had higher formant frequencies following training and sounded more feminine/less masculine to listeners. Similarly, Hirsch[78] presented a resonance training approach developed to modify vocal tract resonance to sound higher or lower/deeper, depending on the goals of the transgender client. The approach uses physiologic behaviors such as lip spreading/rounding and tongue carriage, as described previously, as well as articulatory precision, and end-of-utterance mouth posture to achieve desired acoustic outcomes. Although not yet supported by efficacy research, the approach reportedly is backed by promising anecdotal evidence. The successful outcomes achieved via Hirsch's training program offer preliminary support for the practical application of the present study's vocal tract resonance findings.

Although the composite average formant frequency measure used in the present study was found to be a significant predictor of perceptual ratings, it may be useful to look at the specific vowel formants in the future in order to ascertain whether or not certain vowel formants are more salient than others. This information would be useful in guiding training programs in terms of knowing which behaviors might be most effective in altering vocal masculinity−femininity. Modifying vocal tract resonance already is a common training goal[75] and experts have recommended that it be targeted in conjunction with f_o.[73] The results of this study provide additional support for these practices and recommendations.

Finally, this is the first study to identify SPL as a significant predictor of masculinity−femininity ratings for a group that includes transgender communicators. It extends the limited evidence provided in the cisgender literature by exploring SPL as a function of auditory-perceptual ratings rather than group differences based on gender identity. Auditory-perceptual ratings are more meaningful for informing voice feminization training because of the relevance to the oft-identified outcome of gender attribution. These results show that not only do men speak louder than women[40] but individuals who are perceived to be more masculine speak louder than those who are perceived to be feminine as well. The results of Holmberg et al.[10] suggested such a relationship existed between SPL and gender attribution for transgender communicators but results were limited to only four participants in that study (n = 26).

The identification of SPL as a significant predictor of perceived masculinity−femininity provides preliminary evidence to support more deliberate targeting of vocal loudness in voice feminization training programs as a means of refining gender presentation (eg, somewhat feminine-sounding versus very feminine-sounding). More research is needed to investigate the effect of changing SPL on training outcomes such as gender attribution or masculinity−femininity ratings and to refine training targets. The softest communicator in this study (who also was rated as most feminine) spoke at 69.39 dBSPL (measured at 10 cm from the mouth), more than one SD below the mean of 74.68 (3.50). The loudest communicator was more than two SD above the mean at 82.43 dBSPL. Based on these results, 70 dBSPL may be an appropriate target for achieving a more feminine-sounding voice.

In summary, the results of this research question support the continued focus on modifying f_o and vowel formant frequencies in training and suggest that modifications in SPL may be worthy of more attention. SPL currently is not a common training target.[74]

Acoustic predictors of naturalness ratings

Multiple linear regression results revealed three significant predictors of vocal naturalness ratings: f_o, average formant frequency, and RoS. The variables contributed to the model to a decreasing degree according to the order listed. F_o was negatively related to vocal naturalness ratings whereas both average vowel formant and RoS were positively related to the perceptual rating.

Given that the raters attributed the majority of the communicators (n = 25) as men, it is logical that lower f_o would be associated with higher vocal naturalness ratings overall as raters would be expecting low-pitched voices for speakers who are men. Conversely, high-pitched voices would not be expected for the majority of speakers and may have sounded unusual.

An analogous argument can be used to explain the positive relationship between vocal naturalness ratings and RoS. The studies that have reported gender differences in RoS found that men spoke faster than women.[34,35] It follows that communicators who spoke faster in this study generally would sound more natural since most of them were thought to be men.

It also is possible that RoS is related to speech naturalness in general and not associated with a particular group of speakers. For example, studies from the stuttering and Alternative and Augmentative Communication literature have shown faster RoS to be associated with higher naturalness ratings.[79,80] Unfortunately, this same relationship was not established for fluent speakers.[79] In the present study, the mean RoS was 3.7 syllables/s (SD = 0.56), which is within normal limits for adult speakers during extemporaneous speech.[81] Eight speakers were between one and two SD slower than the group mean and of those, seven spoke at a rate slower than three syllables/s. This rate falls more than two SD below the mean for American English speakers reported in Robb et al[81] Five of the speakers who spoke slower than 3 syllables/s were assigned an average vocal naturalness rating that was below the group mean of 17.64 (1.98). Most of the communicators (n = 25) were rated to be at or above average in terms of naturalness. It stands to reason that extremely slow RoS would sound unusual or unnatural to some listeners and as such, should be avoided by those seeking to develop speech patterns characteristic of a typical speaker. More research is needed to confirm or clarify this relationship for transgender communicators.

The positive relationship between average formant frequency and vocal naturalness ratings is more difficult to interpret than the other two relationships. As average formant frequency increased, so too did the naturalness ratings. This result can be explained on the basis of gender attribution for those who were attributed as women (n = 11) because one would expect to higher vowel formants for feminine-sounding speakers. This argument does not hold, however, for the majority of the communicators who were attributed as men.

Some clarity may be gleaned from Hardy et al,[27] a study reporting similar findings. Those authors examined acoustic predictors of vocal naturalness for a group of transgender women (n = 25) who had not yet received voice and communication feminization training and found F2 of /α/ to be one of three variables included in the predictive model. It was hypothesized that the transgender women communicators were perhaps speaking in a way that resulted in raised F2 values (eg, anterior tongue carriage), which made them feel more feminine and, consequently, more comfortable, which in turn made them sound more natural to listeners. That also may have been the case for the transgender communicators included in the present study, who account for half of the participants. As such, between the two hypotheses, the positive relationship can be explained for most of the study participants.

The results of this research question suggest that f_o and average vowel formant are not only important for modifying the perceived masculinity−femininity of the voice but also are important to ensure the voice sounds natural or authentic. Clinicians also should be cognizant of the influence of RoS on vocal naturalness and discourage clients from speaking at an unusually slow rate.

Relationships between personal factors and auditory-perceptual ratings

Given the broad inclusion criteria for the transgender communicators and the aim of the present research to contribute to the evidence-base informing voice and communication feminization training, the potential impact of transition-related variables is worthy of brief mention. Whereas one might assume that certain of these factors (eg, length of voice and communication training) are related to auditory-perceptual ratings, in fact, there were no systematic relationships between the ratings and time since transition commenced, length of time living fulltime in the feminine gender, length of voice and communication training, time since the completion of, or TVQ^{MtF} total score for the participants in this study. These results are in keeping with previous findings reporting a lack of significant relationships between a variety of personal factors and listener-reported auditory-perceptual ratings of voice for transgender speakers.[10,18] Smoking may have impacted voice quality for some of the communicators; however, it was not formally tested. Although length of training was not significantly related to auditory perceptual ratings in the present study, it is not known whether those who had participated in training had achieved and/or maintained their training goals. Given that mean length of training was so short (ie, approximately 6 weeks), it is possible and perhaps probable that participants had not generalized skills to spontaneous connected speech. In general, the ratings obtained in the present study appeared to be independent of the known personal factors. It is possible that other, unmeasured characteristics may have systematically varied with the ratings.

In summary, the present study successfully identified acoustic predictors of the three auditory-perceptual ratings of interest. F_o was the sole predictor of gender attribution ratings; f_o, average formant frequency, and SPL significantly predicted of masculinity−femininity ratings; and f_o, average formant frequency, and RoS significantly predicted naturalness ratings.

Limitations and future directions

This study was limited by the lack of variability in the transgender participants despite the relatively large sample size. As was described in section 4.1 “Acoustic Predictors of Gender Attribution,” 16/20 transgender participants were attributed as men. Only one was attributed as a woman and three were not attributed consistently as either gender. These results are not completely unexpected given that more than half of these communicators (n = 12) had not yet participated in voice and communication feminization training.

There also may have been a self-selection bias amongst the individuals who volunteered for the study. A concerted effort was made to reach as many transgender women in the community as possible; however, many who were known to have made noticeable changes to their voices (eg, raised f_o) or were successfully “stealth” (self-described) did not volunteer for the study. A small number made inquiries about the line of research but ultimately chose not to participate. Perhaps those who struggled more with their voices and the attributions made by others felt more compelled to contribute to the research whereas those who were having less difficulty had moved forward to “Getting on with Life” or post-transition stage described by Byrne[4] and had less interest in being involved. It is possible that other factors such as restrictions on participation time and location also may have contributed to the recruitment challenges.

Another limitation of the present study was sample size. Although large when compared to previous communication studies involving transgender participants, the sample size was relatively small for the complex statistical tests and number of associated predictor variables used. Results would have been more robust and may have revealed additional significant predictors with a greater sample size: predictor variable ratio, especially in the case of the logistic regression.[82] Similarly, attributed gender groups were unequal, which is not ideal for logistic regression.[82]

In summary, other variables may have emerged as predictors of the auditory-perceptual ratings (especially gender attribution) had there been greater representation of feminine-sounding speakers. Future studies should aim to replicate these results with a larger sample size and more even distribution of transgender participants across the spectrum of gender attribution. As a next step, it would be valuable to test the models created in the present study and any replication studies that follow in terms of their ability to correctly predict auditory-perceptual ratings. If accurate, the associated regression equations potentially could be used as clinical outcome measures. It also would be valuable to explore how these acoustic variables relate to communicator-based outcome measures such as satisfaction with voice.

A final limitation of the present study was the exclusion of certain acoustic variables. As mentioned in section 4.1 "Acoustic Predictors of Gender Attribution," several variables that have evidence to support their contribution to auditory-perceptual ratings related to gender attribution were measured but not included in the regression analyses. In particular, the exclusions unfortunately resulted in a lack of representation of intonation amongst the predictor variables. Future researchers should determine reliable methods of measuring intonation variables such as percentage of upward intonation shifts used and investigate their salience along with those predictors already identified as significant.

The present study also excluded variables related to other aspects of communication such as nonverbal communication (eg, gesture use). These behaviors may be addressed in voice and communication modification training[16] yet there is a dearth of research exploring their role in cueing gender attribution for transgender communicators. Future studies should systematically measure nonverbal communication behaviors and determine whether they are related to perceptual ratings associated with training outcomes.

STUDY CONTRIBUTIONS AND CONCLUSIONS

This study used a unique set of acoustic variables to predict both gender attribution and perceived masculinity−femininity and moved away from the gender binary to include participants whose gender was attributed ambiguously. The inclusion of naturalness ratings provided needed insight into the acoustic factors that contribute to a natural or authentic-sounding voice. Results suggest clinicians should continue to target pitch and vocal tract resonance in their training programs and consider a greater focus on vocal loudness adjustments. By addressing these voice characteristics and avoiding speaking very slowly, transgender clients may be able to achieve a voice that satisfactorily represents their affirmed gender.

Acknowledgments

The authors would like to thank the research participants for their contribution to the research. We also thank the team at Webzao Innovations for their tireless work in developing Gender Finder; Ming Ye for guidance with statistical analyses; Simone Emery, Jessica Frankel, Brianne Haeusler, and Riley Steel for assistance with acoustic analyses; and Matthew Kelley for creating PRAAT scripts for the acoustic analyses. This research was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) and Alberta Innovates.

The funding sources had no involvement in study design; data collection, analysis, or interpretation; writing of the manuscript; or the decision to submit the manuscript for publication.

REFERENCES

1. Stryker S. *Transgender History*. Berkeley, CA: Seal Press; 2008.
2. Hancock A. An ICF perspective on voice-related quality of life of American transgender women. *J Voice*. 2017;31:115e1–115e8. http://dx.doi.org/10.1016/j.jvoice.2016.03.013.
3. Pasricha N, Dacakis G, Oates J. Communicative satisfaction of male-to-female transsexuals. *Logop Phoniatr Vocol*. 2008;33:25–34. http://dx.doi.org/10.1080/14015430701514500.
4. Byrne LA. *My Life as a Woman: Placing Communication Within the Social Context of Life for Transsexual Women, Unpublished Dissertation*. La Trobe University; 2007.
5. Bauer G, Ayden IS. Transgender people in Ontario, Canada: statistics from the Trans PULSE Project to Inform Human Rights Policy, London, ON, 2015.
6. Grant JM, Mottet LA, Tanis J, et al. Injustice at Every Turn: a Report of the National Transgender Discrimination Survey, Washington, DC, 2011.
7. James SE, Herman JL, Rankin S, et al., The Report of the 2015 U.S. Transgender Survey, Washington, DC, 2016.
8. Coleman E, Bockting W, Botzer M, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, Version 7. *Int J Transgenderism*. 2011;13:165–232. http://dx.doi.org/10.1080/15532739.2011.700873.
9. Skuk VG, Schweinberger SR. Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *J Speech, Lang Hear Res*. 2014;57:285–296. http://dx.doi.org/10.1044/1092-4388(2013/12-0314)(VTL).
10. Holmberg EB, Oates J, Dacakis G, et al. Phonetograms, aerodynamic measurements, self-evaluations, and auditory perceptual ratings of male-to-female transsexual voice. *J voice*. 2010;24:511–522. http://dx.doi.org/10.1016/j.jvoice.2009.02.002.
11. Leung Y, Oates J, Pang Chan S. Voice, articulation, and prosody contribute to listener perceptions of speaker gender: a systematic review and meta-analysis. *J speech, Lang Hear Res*. 2018;61:266–297. http://dx.doi.org/10.1044/2017_JSLHR-S-17-0067.
12. Dacakis G, Oates J, Douglas J. Beyond voice: perceptions of gender in male-to-female transsexuals. *Curr Opin Otolaryngol Head Neck Surg*. 2012;20:165–170. http://dx.doi.org/10.1097/MOO.0b013e3283530f85.
13. Gelfer MP, Bennett QE. Speaking fundamental frequency and vowel formant frequencies: effects on perception of gender. *J Voice*. 2013;27:556–566. http://dx.doi.org/10.1016/j.jvoice.2012.11.008.
14. Hillenbrand JM, Clark MJ. The role of f 0 and formant frequencies in distinguishing the voices of men and women. *Atten Percept Psychophys*. 2009;71:1150–1166. http://dx.doi.org/10.3758/APP.71.5.1150.
15. Oates J, Dacakis G. Transgender voice and communication: research evidence underpinning voice intervention for male-to-female transsexual women. *Perspect Voice Voice Disord*. 2015;25:48–58. http://dx.doi.org/10.1044/vvd25.2.48.
16. Adler R, Hirsch S, Mordaunt M, eds. *Voice and Communication Therapy for the Transgender/Transsexual Client: A Comprehensive Clinical Guide*. San Diego: Plural Publishing Inc.; 2006.
17. Hancock A, Colton L, Douglas F. Intonation and gender perception: applications for transgender speakers. *J Voice*. 2014;28:203–209. http://dx.doi.org/10.1016/j.jvoice.2013.08.009.
18. Owen K, Hancock AB. The role of self- and listener perceptions of femininity in voice therapy. *Int J Transgenderism*. 2010;12:272–284. http://dx.doi.org/10.1080/15532739.2010.550767.
19. Gelfer MP, Mikos VA. The relative contributions of speaking fundamental frequency and formant frequencies to gender identification

based on isolated vowels. *J Voice*. 2005;19:544–554. http://dx.doi.org/10.1016/j.jvoice.2004.10.006.
20. Gelfer MP, Tice RM. Perceptual and acoustic outcomes of voice therapy for male-to-female transgender individuals immediately after therapy and 15 months later. *J Voice*. 2013;27:335–347. http://dx.doi.org/10.1016/j.jvoice.2012.07.009.
21. McNeill EJM, Wilson JA, Clark S, et al. Perception of voice in the transgender client. *J Voice*. 2008;22:727–733. http://dx.doi.org/10.1016/j.jvoice.2006.12.010.
22. Oates JM, Dacakis G. Voice change in transsexuals. *Venereology*. 1997;10:178–187.
23. Van Borsel J, De Cuypere G, Van den Berghe H. Physical appearance and voice in male-to-female transsexuals. *J Voice*. 2001;15:570–575. http://dx.doi.org/10.1016/S0892-1997(01)00059-5.
24. Wolfe V, Ratusnik D, Smith F, et al. Intonation and fundamental frequency in male-to-female transsexuals. *J Speech Hear Disord*. 1990;55:43–50.
25. Carew L, Dacakis G, Oates J. The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *J Voice*. 2007;21:591–603. http://dx.doi.org/10.1016/j.jvoice.2006.05.005.
26. Gorham-Rowan M, Morris R. Aerodynamic analysis of male-to-female transgender voice. *J Voice*. 2006;20:251–262. http://dx.doi.org/10.1016/j.jvoice.2005.03.004.
27. Hardy TLD, Boliek CA, Wells K, et al. Pretreatment acoustic predictors of gender, femininity, and naturalness ratings in individuals with male-to-female gender identity. *Am J Speech-Lang Pathol*. 2016;25:125–137. http://dx.doi.org/10.1044/2015_AJSLP-14-0098.
28. Mordaunt M. Pitch and intonation. In: Adler RK, Hirsch S, Mordaunt M, eds. *Voice and Communication Therapy for the Transgender/Transsexual Client: A Comprehensive Clinical Guide*. San Diego, CA: Plural Publishing Inc.; 2006:168–208.
29. Gelfer MP, Mordaunt M. Pitch and Intonation. In: Adler RK, Hirsch S, Mordaunt M, eds. *Voice and Communication Therapy for the Transgender/Transsexual Client: A Comprehensive Clinical Guide*. second ed. San Diego, CA: Plural Publishing Inc.; 2012:187–224.
30. Günzburger D. Acoustic and perceptual implications of the transsexual voice. *Arch Sex Behav*. 1995;24:339–348.
31. Mount KH, Salmon SJ. Changing the vocal characteristics of a postoperative transsexual patient: a longitudinal study; *J Commun Disord*. 1988;21:229–238. http://www.ncbi.nlm.nih.gov/pubmed/3417881.
32. Gelfer MP, Schofield KJ. Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male; *J Voice*. 2000;14:22–33. http://www.ncbi.nlm.nih.gov/pubmed/10764114.
33. King RS, Brown GR, McCrea CR. Voice parameters that result in identification or misidentification of biological gender in male-to-female transgender veterans. *Int J Transgenderism*. 2012;13:117–130. http://dx.doi.org/10.1080/15532739.2011.664464.
34. Verhoeven J, De Pauw G, Kloots H. Speech rate in a pluricentric language: a comparison between Dutch in Belgium and the Netherlands. *Lang Speech*. 2004;47:297–308. http://dx.doi.org/10.1177/00238309040470030401.
35. Fitzsimons M, Sheahan N, Staunton H. Gender and the integration of acoustic dimensions of prosody: implications for clinical studies. *Brain Lang*. 2001;78:94–108. http://dx.doi.org/10.1006/brln.2000.2448.
36. Tsao YC, Weismer G. Interspeaker variation in habitual speaking rate: evidence for a neuromuscular component. *J Speech Hear Res*. 1997;40:858–866.
37. Tsao Y, Weismer G, Iqbal K. Interspeaker variation in habitual speaking rate : additional evidence. *J Speech, Lang Hear Res*. 2006;49:1156–1165.
38. Van Borsel J, De Maesschalck D. Speech rate in males, females, and male-to-female transsexuals. *Clin Linguist Phonetics*. 2008;22:679–685. http://dx.doi.org/10.1080/02699200801976695.
39. Günzburger D. Voice adaptation by transsexuals. *Clin Linguist Phonetics*. 1989;3:163–172.
40. Gelfer MP, Young SR. Comparisons of intensity measures and their stability in male and female speakers. *J Voice*. 1997;11:178–186.
41. Boonin J. Rate and volume. In: Adler RK, Hirsch S, Mordaunt M, eds. *Voice and Communication Therapy for the Transgender/Transsexual Client: A Comprehensive Clinical Guide*. second ed.San Diego, CA: Plural Publishing Inc.; 2012:237–251.
42. Oates JM, Dacakis G. Speech pathology considerations in the management of transsexualism—a review. *Br J Disord Commun*. 1983;18:139–151. http://dx.doi.org/10.3109/13682828309012237.
43. Günzburger D. An acoustic analysis and some perceptual data concerning voice change in male-female trans-sexuals. *Eur J Disord Commun*. 1993;28:13–21.
44. Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers; *J Acoust Soc Am*. 1990;87:820–857. http://www.ncbi.nlm.nih.gov/pubmed/2137837.
45. Van Borsel J, Janssens J, De Bodt M. Breathiness as a feminine voice characteristic: a perceptual approach. *J Voice*. 2009;23:291–294. http://dx.doi.org/10.1016/j.jvoice.2007.08.002.
46. Södersten M, Hertegård S, Hammarberg B. Glottal closure, transglottal airflow, and voice quality in healthy middle-aged women; *J Voice*. 1995;9:182–197. http://www.ncbi.nlm.nih.gov/pubmed/7620541.
47. Mendoza E, Valencia N, Muñoz J, et al. Differences in voice quality between men and women: use of the long-term average spectrum (LTAS); *J Voice*. 1996;10:59–66. http://www.ncbi.nlm.nih.gov/pubmed/8653179.
48. Andrews ML, Schmidt CP. Gender presentation: perceptual and acoustical analyses of voice. *J Voice*. 1997;11:307–313.
49. Wendahl RW. Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. *Folia Phoniatr (Basel)*. 1966;18:98–108.
50. Wendahl RW. Some parameters of auditory roughness. *Folia Phoniatr Logop*. 1966;18:26–32.
51. Bhuta T, Patrick L, Garnett JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice*. 2004;18:299–304. http://dx.doi.org/10.1016/j.jvoice.2003.12.004.
52. Porter CC. *Voice quality and gender identification: acoustic and perceptual analysis*, Unpublished Master's Thesis. Dalhousie University; 2012.
53. Dacakis G, Davies S. Transsexual voice questionnaire (male-to-female). 2012.
54. Martins RHG, Tavares ELM, Pessin ABB. Are vocal alterations caused by smoking in Reinke's Edema in women entirely reversible after microsurgery and smoking cessation. *J Voice*. 2017;31:380.e11–380.e14. http://dx.doi.org/10.1016/j.jvoice.2016.06.012.
55. Dacakis G, Oates J, Douglas J. Associations between the transsexual voice questionnaire (TVQ MtF) and self-report of voice femininity and acoustic voice measures. *Int J Lang Commun Disord*. 2017;52:831–838. http://dx.doi.org/10.1111/1460-6984.12319.
56. Hancock AB, Pool SF. Influence of listener characteristics on perceptions of sex and gender. *J Lang Soc Psychol*. 2017;36:599–610. http://dx.doi.org/10.1177/0261927X17704460.
57. Milenkovic PH. TF32 [Computer program]. 2001.
58. Davis A. *In the Pink of the Night*. Metro-Goldwyn-Mayer; 1969.
59. Baken R, Orlikoff R. *Clinical Measurement of Speech and Voice*. second ed San Diego, CA: Singular Publishing Group; 2000.
60. Fox CM, Boliek CA. Intensive voice treatment (LSVT LOUD) for children with spastic cerebral palsy and dysarthria. *J Speech, Lang Hear Res*. 2012;55:930–945. http://dx.doi.org/10.1044/1092-4388(2011/10-0235).
61. Hunt KW. A synopsis of clause-to-sentence length factors. *English J*. 1965;54:305–309.
62. Boersma P, Weenink D. PRAAT: doing phonetics by computer [Computer program] Version 6.0.26; 2017. Published; 2017. http://www.praat.org/. Published.
63. Kelley M. Boliek lab segmental variables extractor [PRAAT script], 2016.
64. Kelley M. Fo Script [PRAAT script], 2017.
65. KayPentax. Multi-Speech SoftwareTM (Model 3700) [Computer program], 2008.

66. Hardy TLD, Rieger JM. *Gender Finder [Computer Program]*. Webzao Innovations Inc.; 2017.
67. Kreiman J, Gerratt BR, Kempster GB, et al. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research; *J Speech Hear Res*. 1993;36:21–40. http://www.ncbi.nlm.nih.gov/pubmed/8450660.
68. Whitehill TL, Lee ASY, Chun JC. Direct magnitude estimation and interval scaling of hypernasality. *J Speech, Lang Hear Res*. 2002;45:80–88. http://dx.doi.org/10.1044/1092-4388(2002/006).
69. Snow MP, Williges RC. Empirical models based on free-modulus magnitude estimation of perceived presence in virtual environments. *Hum Factors Ergon Soc*. 1998;40:1–14.
70. Hosmer DJ, Lemeshow S, Sturdivant R. *Model-Building Strategies and Methods for Logistic Regression, in: Applied Logistic Regression*. third ed. New Jersey: John Wiley & Sons, Inc; 2013:89–151.
71. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropractic Med*. 2016;15:155–163. http://dx.doi.org/10.1016/j.jcm.2016.02.012.
72. Spencer LE. Speech characteristics of male-to-female transsexuals: a perceptual and acoustic study. *Folia Phoniatr (Basel)*. 1988;40:31–42.
73. 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. http://dx.doi.org/10.1080/15532739.2015.1075931.
74. Hancock AB, Garabedian LM. Transgender voice and communication treatment: a retrospective chart review of 25 cases. *Int J Lang Commun Disord*. 2013;48:54–65. http://dx.doi.org/10.1111/j.1460-6984.2012.00185.x.
75. Dahl K. *Correlating speech and voice features of transgender women with ratings of femininity and gender*; . unpublished master's thesis. University of Rhode Island; 2018. http://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=2229&context=theses.
76. Coleman RO. A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *J Speech Hear Res*. 1976;19:168–180.
77. Gelfer MP, Van Dong BR. A preliminary study on the use of vocal function exercises to improve voice in male-to-female transgender clients. *J Voice*. 2013;27:321–334. http://dx.doi.org/10.1016/j.jvoice.2012.07.008.
78. Hirsch S. Combining voice, speech science and art approaches to resonant challenges in transgender voice and communication training. *Perspect ASHA Special Interest Groups SIG 10*. 2017;2:74–83.
79. Van Borsel J, Eeckhout H. The speech naturalness of people who stutter speaking under delayed auditory feedback as perceived by different groups of listeners. *J Fluency Disord*. 2008;33:241–251. http://dx.doi.org/10.1016/j.jfludis.2008.06.004.
80. Ratcliff A, Coughlin S, Lehman M. Factors influencing ratings of speech naturalness in augmentative and alternative communication. *AAC Augment Altern Commun*. 2002;19:11–19.
81. Robb MP, Maclagan MA, Chen Y, et al. Speaking rates of American and New Zealand varieties of English. *Clin Linguist Phon*. 2004;18:1–15. http://dx.doi.org/10.1080/0269920031000105336.
82. Tabachnick B, Fidell L. *Using Multivariate Statistics*. Boston: Pearson; 2013.