Interspeech 2022
18-22 September 2022, Incheon, Korea

Effects of laryngeal manipulations on voice gender perception

Zhaoyan Zhang[1], *Jason Zhang*[1], *Jody Kreiman*[1]

[1]Department of Head and Neck Surgery, University of California, Los Angles
zyzhang@ucla.edu, jzhang17@geffenacademy.ucla.edu, jkreiman@ucla.edu

Abstract

This study aims to identify laryngeal manipulations that would allow a male to approximate a female-sounding voice, and that can be targeted in voice feminization surgery or therapy. Synthetic voices were generated using a three-dimensional vocal fold model with parametric variations in vocal fold geometry, stiffness, adduction, and subglottal pressure. The vocal tract was kept constant in order to focus on the contribution of laryngeal manipulations. Listening subjects were asked to judge if a voice sounded male or female, or if they were unsure. Results showed the expected large effect of the fundamental frequency (F0) and a moderate effect of spectral shape on gender perception. A mismatch between F0 and spectral shape cues (e.g., low F0 paired with high H1-H2) contributed to ambiguity in gender perception, particularly for voices with F0 in the intermediate range between those of typical adult males and females. Physiologically, the results showed that a female-sounding voice can be produced by decreasing vocal fold thickness and increasing vocal fold transverse stiffness in the coronal plane, changes in which modified both F0 and spectral shape. In contrast, laryngeal manipulations with limited impact on F0 or spectral shape were shown to be less effective in modifying gender perception.

Index Terms: gender perception, transgender voice, laryngeal manipulation, vocal fold thickness

1. Introduction

Gender recognition from the voice is usually easy and often accurate [1]. However, most people can imitate voices of both males and females. In the clinic, voice therapy or surgery is often used to alter voice production in transgender speakers to match their gender identity [2-5]. While there have been many studies on identifying acoustic markers of gender perception [1], few studies investigated physiological manipulations that produce voices of a desired gender identity and their effectiveness in achieving this goal. The purpose of our study is to identify such physiological manipulations and evaluate their effectiveness in altering gender perception. While many aspects of voice production need to be modified in order to successfully conform to the desired gender identity, in this study we focus on the effect of laryngeal manipulations on voice gender perception.

A better understanding of the effect of laryngeal manipulations on gender perception is of particular clinical importance. There has been an increasing interest in improving clinical management of transgender voice [2, 3, 6, 7]. While gender perception depends on many factors including for example pitch, vocal tract resonance, articulation, and speaking style, modification of the fundamental frequency (F0) often remains the primary target of clinical intervention [3]. This is particularly the case for voice feminization surgery, which aims to modify the length, stiffness, and mass of male vocal folds in order to increase the produced F0 toward the adult female range. Different surgical techniques have been developed [8]. However, despite many voice outcome studies comparing different techniques, there have been few systematic investigations on how such surgical procedures impact voice production mechanisms [9] and their effectiveness at modification of gender perception toward the desired gender identity.

In general, adult male vocal folds are longer and thicker than adult female vocal folds [10, 11]. The length difference is generally considered responsible for the large difference in fundamental frequency between adult males and females [10], which plays an important role in gender perception. As a result, voice feminization surgery often aims to reduce vocal fold length and/or increase vocal fold longitudinal stiffness in order to increase the produced F0. However, increasing mean F0 to within the female range does not necessarily result in a voice that is perceived as female [2]. On the other hand, the thickness difference between adult male and female vocal folds has been shown to be largely responsible for the often-reported differences between adult males and females in the glottal closure pattern and the resulting spectral shape, another important contributor to gender perception [12]. Thus, modification of vocal fold thickness, which simultaneously modifies F0 and spectral shape, is likely to have a large impact on gender perception.

In this study, using a computational voice production model, we performed voice production simulations with parametric variations in vocal fold thickness, stiffness, adduction, and subglottal pressure. The perceived gender of the produced voices was then evaluated in a listening experiment. We will show that laryngeal manipulations that modify F0 and spectral shape simultaneously are more effective in modifying gender perception and avoiding gender ambiguity than manipulations that impact mostly F0 alone.

2. Methods

2.1. Synthetic voices

Synthetic voices were generated in voice production simulations using a three-dimensional vocal fold model [13, 14, 15] (Figure 1) with parametric variations in vocal fold geometry, stiffness, and position. In this study, parametric variations in the initial glottal angle (a measure of the degree of vocal fold adduction), vertical thickness of the vocal fold medial surface, vocal fold stiffness along the longitudinal (anterior-posterior) direction, vocal fold transverse stiffness in the coronal plane, and subglottal pressure were considered, as shown in Table 1.

In order to focus on laryngeal manipulations that can be made in a male larynx to produce a female-sounding voice, in

 10.21437/Interspeech.2022-10815

this study the vocal fold length was kept at 17 mm, a typical value for membranous vocal fold length in males, for all simulations. A constant male vocal tract was used, with the shape corresponding to the /ɑ/ sound [16], also in order to focus on laryngeal manipulations. For each vocal fold condition and a given subglottal pressure, a sustained /ɑ/ was simulated for 0.5 seconds, as in our previous studies [13-15].

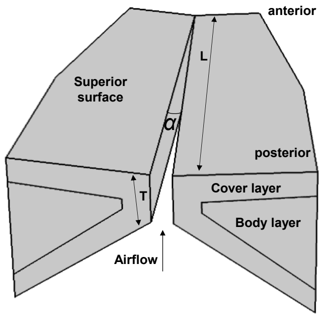

Figure 1: *The three-dimensional vocal fold model and key geometric control parameters, including the vocal fold length L along the anterior-posterior direction, vertical thickness of the medial surface T, and the initial glottal angle a.*

Table 1: *Ranges of parametric variations used in voice simulations to generate the synthetic voices.*

Parameter	Values
initial glottal angle α (°)	0, 1.6, 4
vertical thickness T (mm)	1, 2, 3, 4.5
cover-layer longitudinal stiffness G_{apc} (kPa)	1, 10, 20, 30, 40
body-layer longitudinal stiffness G_{apb} (kPa)	1, 10, 20, 30, 40
transverse stiffness Et (kPa)	1, 2, 4
subglottal pressure Ps (Pa)	50 – 2400

2.2. Listening experiment

A set of 1000 voices were randomly selected from the simulated voices and used in a listening experiment. The voices were selected so that each of the six control parameters was approximately evenly distributed in their ranges listed in Table 1. Voices were normalized for intensity. Eleven subjects (6 females; average age of 26 years old) participated in the listening experiment. The participants listened to the voices over headphones, and were asked to decide for each voice whether it sounded more like a male or a female, or they were not sure. Given that some of the synthetic voices may sound pathological, subjects were further instructed to focus on the gender (male or female) of the voices rather than other aspects of the voice quality. Subjects could play each voice up to five times. They were asked to take breaks when necessary. Each experiment lasted less than one hour.

2.3. Data analysis

For each voice and each subject, a gender score was assigned based on subject response: 0 for a male response, 1 for unsure, and 2 for a female response. Listeners showed moderate agreement in their responses. A mean gender score was calculated for each voice by averaging the gender scores from all subjects. Acoustic measures were also extracted for each voice, including the fundamental frequency (F0), cepstral peak prominence (CPP), amplitude differences between the first harmonic and the second harmonic (H1-H2), the fourth harmonic (H1-H4), the harmonic nearest 2 kHz (H1-H2k), and the harmonic nearest 5 kHz (H1-H5k) in the spectrum of the time derivative of the glottal flow waveform. These measures have been shown to be perceptually important [17]. The mean gender score was correlated with the acoustic measures and physiological controls using multiple linear regression. Analysis of variance (ANOVA) was then performed to investigate the effect sizes of the six physiological controls on the mean gender score.

To better characterize acoustic differences contributing to differences in gender perception, three subcategories of the synthetic voices were generated. The first category, labeled 'male' in the following, includes voices with a mean gender score below 0.25. The second category, labeled 'unsure', includes voices with a mean gender score between 0.75 and 1.25. The last category, labeled 'female', includes voices with a mean gender score above 1.75. These three voice categories were intentionally separated in terms of the mean gender score to increase acoustic contrast between these categories, thus facilitating identification of acoustic characteristics contributing to differences in gender perception.

3. Results

3.1. Correlation between gender perception and acoustics

Table 2 shows the results from multiple linear regression between the mean gender scores and the acoustic measures. F0 had the largest standard coefficient, followed by H1-H2 and H1-H5k. All three measures were statistically significant. The dominant effect of F0 on the mean gender score is consistent with findings from previous studies [1]. Weak correlations were also observed for the other acoustic measures, but were not statistically significant.

Table 2: *Standard coefficients of linear regression between mean gender score and acoustic measures for all voices.*

Acoustics	Standard coefficient	p value
F0	**0.753**	**<0.005**
H1-H2	**0.114**	**<0.005**
H1-H4	-0.062	0.030
H1-H2k	0.056	0.135
H1-H5k	**-0.110**	**<0.005**
CPP	-0.006	0.780

To better understand why subjects were unsure about the gender of some voices, three voice categories (male, female, unsure) were generated, as described above in the Methods section. Figure 2 shows the F0 distributions of the three categories. In general, voices in the male category had a low F0 in the adult male range, whereas voices in the female category had a high F0 in the adult female range. Voice in the unsure category had a F0 in the intermediate range. It is interesting to note that some of the unsure voices had F0 well inside the typical range of adult female voices, yet still received an unsure response.

The three voice categories (male, female, unsure; see Methods section) were next projected onto the acoustic space identified by principal component analysis (PCA) applied to all the synthetic voices. Figure 3a shows the first three most dominant PCA modes. In this study, the PCA modes were interpreted by only considering variables with loadings of 0.3 and above. Thus, the first PCA mode represents covariations in the four spectral slope measures. PCA-2 represents covariations between F0 and the spectral shape measures H1-H2, H1-H2k, and H1-H5k. PCA-3 is dominated by CPP, a measure of the relative strength between harmonics and noise. The same acoustic variables also emerge in PCA modes of natural human voices [18], confirming the validity of the synthetic voices for perceptual studies.

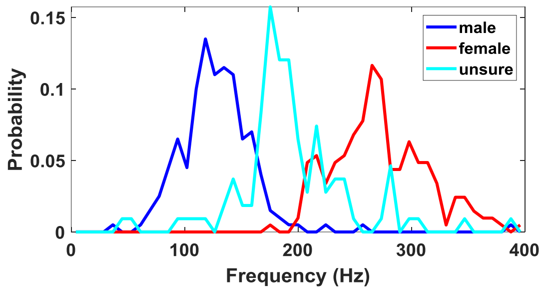

Figure 2: *Fo distribution of the male, female, and unsure voice categories.*

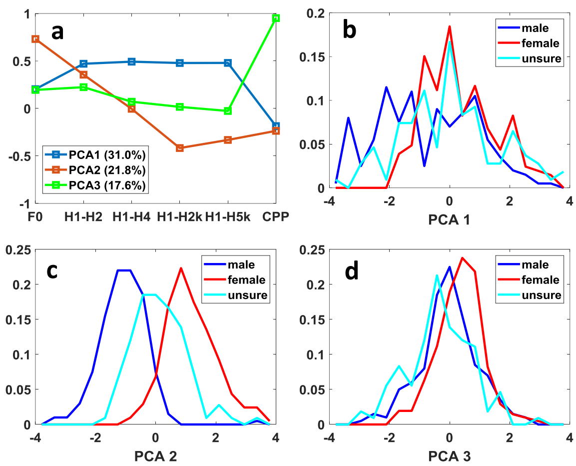

Figure 3: *The first three PCA modes of the acoustic space, and the representations of three voice categories (male, female, unsure) in the PCA space. The numbers in parentheses show the percentage of variance explained by each PCA modes.*

Figures 3b-3d show the projections of the three voice categories on the first three PCA modes. A clear difference between the three voice categories can be observed in their projections on the PCA-2 mode, whereas the distributions of their projections onto the other two modes were relatively similar. In other words, the three voice categories differed from each other in terms of the covariations, or lack of covariations, between F0 and spectral shape. For both the male and female voice categories, there was a strong covariation between F0 and spectral shape: the male voice category was characterized by low F0 and low H1-H2, whereas the female voice category was characterized by a high F0 and a high H1-H2. Such covariations between F0 and spectral shape was weaker in the unsure voice category, as demonstrated by the weak representation in the PCA-2 mode (i.e., coefficients close to zero in figure 3c). For example, some voices in the unsure category had a pitch close to the typical adult male range but an H1-H2 value that was high in the female range. This mismatch in F0 and spectral shape cues may have contributed to the unsure response from the subjects.

Table 3: *Standard coefficients of linear regression between mean gender score and acoustic measures for voices in the male and female categories only.*

Acoustics	Standard coefficient	p value
F0	**0.894**	**<0.005**
H1-H2	0.032	0.434
H1-H4	-0.038	0.279
H1-H2k	0.082	0.045
H1-H5k	-0.076	0.034
CPP	0.018	0.447

If the mismatch in F0 and spectral shape cues was indeed the cause of ambiguity in gender perception in the unsure voice category, we would expect that spectral shape measures are less important and F0 plays a dominant role in gender perception in the male and female voice categories, in which the spectral shape cues are consistent with the F0 cue. This was confirmed in Table 3, which shows the linear regression between the mean gender score and acoustic measures for voices in the male and female categories, excluding the unsure category. In this case, F0 became the only predictor of the mean gender score in Table 3 to achieve statistical significance.

3.2. Correlation between gender perception and physiology

Table 4 shows the results of multiple linear regression and analysis of variance (ANOVA) of the mean gender score as a function of the six physiological controls. Figure 4 further shows the general trends of variation of the mean gender score with each physiological control parameter (except subglottal pressure, for which no obvious trend could be identified).

Table 4: *Standard coefficients, F value, and effect size η^2 between the mean gender score and physiological controls.*

Physiology	Standard coefficient	F/η^2
Subglottal pressure Ps	0.207	6/0.064
Vertical thickness T	**-0.382**	**81/0.147**
Initial glottal angle α	-0.249	58/0.071
Transverse stiffness Et	**0.328**	**87/0.105**
Cover-layer longitudinal stiffness G_{apc}	0.161	13/0.033
Body-layer longitudinal stiffness G_{apb}	0.070	4/0.009

Vertical thickness had the largest effect size on the mean gender score, followed by transverse stiffness (Table 4, third column). Both had an effect size above 0.1. The mean gender score increased (i.e., more female-sounding) with decreasing thickness or increasing transverse stiffness (Table 4, 2nd column). A moderate effect on the mean gender score was observed for the initial glottal angle, with the two smaller glottal angles producing significantly more female-sounding voices than the largest glottal angle (figure 4). A similar moderate effect was also observed for subglottal pressure and cover-layer longitudinal stiffness. Body-layer longitudinal stiffness had the smallest effect size.

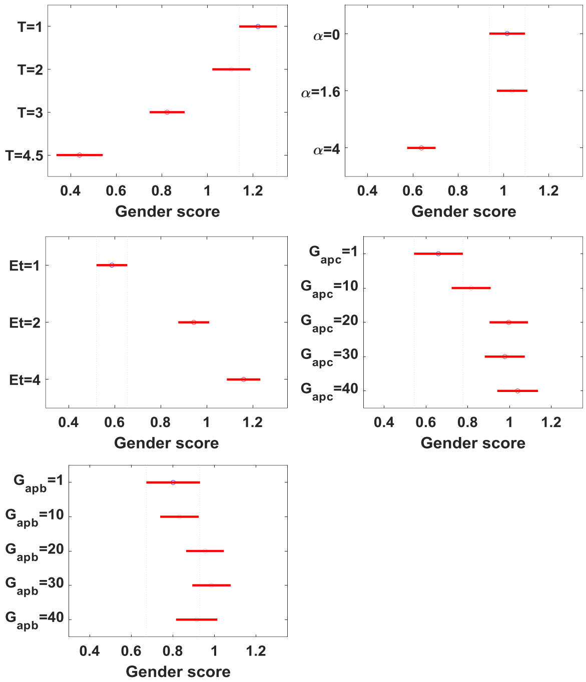

Figure 4: *Trends of variation of the mean gender score with the five physiological controls.*

Table 5: *F value/effect size η^2 of the physiological controls on F0 and H1-H2.*

Physiology	F0	H1-H2
Subglottal pressure	6/0.056	2/0.029
Vertical thickness	36/0.064	**74/0.181**
Initial glottal angle	82/0.097	0/0.000
Transverse stiffness	**116/0.138**	53/0.086
cover-layer longitudinal stiffness	24/0.058	3/0.010
Body-layer longitudinal stiffness	8/0.019	5/0.015

Table 5 shows the F values and effect sizes of the physiological controls on F0 and H1-H2. Comparison between Tables 4 and 5 further supports our earlier observation that F0 is not the only parameter determining gender perception. For example, vertical thickness had the largest effect on the mean gender score, despite only a moderate effect on F0. The large effect of vertical thickness on gender perception was likely related to its large effect on the spectral shape (e.g., H1-H2 in Table 5), which may have facilitated gender perception. In contrast, the initial glottal angle had a relatively large effect on F0, but a very small effect on H1-H2, which may have contributed to the relatively small effect of the initial glottal angle on the mean gender score.

4. Discussion

Our results showed that in addition to F0, spectral shape of the voice source (H1-H2, H1-H5k in this study) also plays an important role in gender perception. A mismatch in F0 and spectral shape may lead to ambiguity in gender perception, particularly when the F0 falls in the intermediate frequency range between typical males and females.

Similarly, laryngeal manipulations that simultaneously modulate both F0 and spectral shape generally had the largest effects on gender perception. Both vertical thickness and transverse stiffness have been shown to have a large effect on the glottal closure pattern and thus on the spectral shape of the voice source [13, 14, 15]. Specifically, reducing vocal fold thickness or increasing vocal fold transverse stiffness increased both F0 and H1-H2, thus producing a more female-sounding voice. This is consistent with the finding that reduced closed quotients, which are correlated with increased H1-H2, have been reported in transgender females in an endoscopic study [19]. On the other hand, changes in the initial glottal angle and vocal fold cover-layer longitudinal stiffness had a moderate to large effect on F0 but only a small effect on the spectral shape, and thus had smaller effects on gender perception.

Our results showed that the two smaller initial glottal angles produced a more female-sounding voice quality. This is largely due to the F0-increasing effect of reducing glottal angle [15]. It should be noted that the three-dimensional vocal fold model currently does not include a posterior cartilaginous opening, which is often present in females and produces a breathy voice quality that is often associated with a female voice. It is possible that when a cartilaginous opening is added to the model, larger glottal angles may increase noise production and thus be more likely to produce a female-sounding voice.

The results suggest that clinical intervention targeting the voice source of transgender female speakers, through either voice therapy or surgery, should focus on laryngeal manipulations that modify F0 and spectral shape simultaneously in order to avoid mismatches. This can be achieved by guiding speakers or through surgery to reduce vocal fold thickness and increase vocal fold transverse stiffness. The results of this study also imply that if clinical intervention results in F0 in the intermediate range between adult males and females, as is often the case, the resulting voice may be perceptually ambiguous if the F0 increase was achieved without notable modification in the spectral shape. This hypothesis is worth further investigating in future studies.

5. Conclusions

Our study showed that both F0 and spectral shape play a role in gender perception. The role of spectral shape is particularly important for voices with F0 in the intermediate range between those of typical males and females, in which a mismatch between the F0 and spectral shape cues may result in ambiguity in gender perception. Thus, laryngeal manipulations that simultaneously modify F0 and spectral shape are more effective in affecting gender perception and avoiding gender perception ambiguity. In this study, these manipulations include reducing vocal fold thickness and increasing vocal fold transverse stiffness.

6. Acknowledgements

This research was supported by research grants R01DC001797 and R01DC009229 from the National Institute on Deafness and other Communication Disorders.

7. References

[1] J. Kreiman, D. Sidtis, *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons, 2011.

[2] G. Dacakis, "The role of voice therapy in male-to-female transsexuals." Current Opinion in Otolaryngology & Head and Neck Surgery 10, no. 3, 173-177, 2002.

[3] G. Dacakis, J. Oates, J. Douglas. "Beyond voice: perceptions of gender in male-to-female transsexuals." Current opinion in otolaryngology & head and neck surgery 20, no. 3, 165-170, 2012.

[4] V. G. Skuk, S. R. Schweinberger, "Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender." Journal of Speech, Language, and Hearing Research 57, 285–296, 2014.

[5] I. T. Nolan, S. Morrison, O. Arowojolu, C. Crowe, J. Massie, R. Adler, S. Chaiet, D. Francis. "The role of voice therapy and phonosurgery in transgender vocal feminization." Journal of Craniofacial Surgery 30, no. 5, 1368-1375, 2019.

[6] L. Zimman, "Transgender voices: Insights on identity, embodiment, and the gender of the voice." Language and Linguistics Compass 12(8), e12284, 2018.

[7] M. Gray, M. Courey. "Transgender voice and communication." Otolaryngologic Clinics of North America 52, no. 4, 713-722, 2019.

[8] H. Kim, "Vocal feminization for transgender women: Current strategies and patient perspectives." International journal of general medicine 13, 43-52, 2020.

[9] I.R. Titze, A. Palaparthi, T. Mau. "Vocal tradeoffs in anterior glottoplasty for voice feminization." The Laryngoscope 131, no. 5, 1081-1087, 2021.

[10] I. R. Titze, "Physiologic and acoustic differences between male and female voices." The Journal of the Acoustical Society of America 85, no. 4, 1699-1707, 1989.

[11] Z. Zhang, "Mechanics of human voice production and control." The Journal of the Acoustical Society of America 140, no. 4, 2614-2635, 2016.

[12] C.R. Pernet, P. Belin. "The role of pitch and timbre in voice gender categorization." Frontiers in psychology. 3, 23, 2012.

[13] Z. Zhang, "Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model." The Journal of the Acoustical Society of America 139, no. 4, 1493-1507, 2016.

[14] Z. Zhang, "Effect of vocal fold stiffness on voice production in a three-dimensional body-cover phonation model." The Journal of the Acoustical Society of America 142, no. 4, 2311–2321, 2017.

[15] Z. Zhang, "Contribution of laryngeal size to differences between male and female voice production." The Journal of the Acoustical Society of America 150, no. 6, 4511-4521, 2021.

[16] B.H. Story, I.R. Titze, E.A. Hoffman, "Vocal tract area functions from magnetic resonance imaging." The Journal of the Acoustical Society of America 100, 537–554, 1996.

[17] J. Kreiman, B. Gerratt, M. Garellek, R. Samlan, Z. Zhang. "Toward a unified theory of voice production and perception." Loquens 1, no. 1, e009, 2014.

[18] Y. Lee, P Keating, and J. Kreiman, "Acoustic voice variation within and between speakers," J. Acoust. Soc. Am. vol 146, pp. 1568, 2019.

[19] D.A. Palmer, Laryngeal structure and function in male-to-female transsexual persons: Endoscopic and stroboscopic presentation. Master's Thesis. University of Kansas, 2009.