Gender Perception After Raising Vowel Fundamental and Formant Frequencies: Considerations for Oral Resonance Research

***Sally J.K. Gallena, *,†Betsy Stickels, and ‡Emily Stickels,** **Baltimore, †Rockville, and ‡Germantown, Maryland*

Summary: Objectives. Voice feminization therapy for male-to-female transgender women typically targets increasing fundamental frequency (F_0). Increasing vowel formant frequencies (FFs) has also been proposed. To better understand formant conditions that shift listeners' perception of gender from male to "not-male," individual and combined vowel FFs were incrementally raised, whereas F_0 was held constant at a gender-ambiguous level.

Methods. The study used a prospective, experimental group design. Using a customized *MATLAB* program, vowels (/i/, /æ/, /ɑ/, and /u/) spoken by an adult were manipulated by isolating and increasing $FF_{1–3}$ until they matched those of a woman. Listeners heard randomized samples and perceptually categorized each as male, female, or gender neutral. The latter two choices were combined and labeled not-male.

Results. Chi-square analyses revealed that listeners rated samples as not-male for /ɑ/ and /æ/ with all three formants shifted or individual formants shifted at >60%. Individual analysis of vowels, formants, and shifted FF using Kruskal-Wallis revealed a statistical significance for vowels only.

Conclusions. Results suggest that voice was convincingly perceived as not-male, for vowels characterized by a high F_1 frequency, and that raising FFs for all four vowels increased (in varying amounts) the perception of voice femininity beyond that of raising F_0 alone.

Key Words: Gender perception–Transgender voice–Signal processing–Formant manipulation–Speech therapy.

During verbal communication, the voice reveals clues about the speaker's gender, age, emotion, and health through variations in pitch, loudness, and quality.[1] Most listeners are adept at perceptually differentiating men from women based solely on the speaker's voice.[2,3] Anatomic differences that account for each gender's unique voice include vocal fold length and mass, which determine the fundamental frequency (F_0) (ie, modal pitch), and the size and the shape of the vocal tract, which determine the resonant frequencies of formants (ie, overtones).[4,5] The average woman's speaking F_0 is approximately 1.7 times higher and her vowel formant frequencies (FFs) are approximately 1.17 times higher than those of an adult.[4,5] These are important anatomic and acoustic differences when considering the voices of individuals who are transgender (TG)—persons who seek to transition from their biological gender to the gender with which they identify.[6] Transgender women (TW) (transitioning from man to woman) are challenged by the need to acquire and to habituate a perceptually feminine voice, in part because of the size of their larynx, specifically the vocal folds, and their vocal tract.[4,5] The incongruence that TW experience between their physical appearance and their voice, at the extreme, threatens their personal safety and job security, whereas daily it undermines their gender identity. As noted by Davies et al,[7] "speech and voice feminization is widely recognized as a vital component of care for TW" (p. 121). Evidence-based speech and voice treatment techniques and outcomes based on quality research designs are lacking for individuals who are TG. The American Speech-Language-Hearing Association's Evidence Maps (http://www.asha.org/Evidence-Maps/), articles by Davies et al[7] and Oates and Dacakis,[8] and a text by Adler et al[9] provide treatment guidelines that need scientific support so that individuals who are TG can experience a successful communication transition.

Current practices in voice therapy for TWs rely heavily on raising the habitual speaking pitch to a level that is no longer perceived as male[8,9] based upon prior research that has indicated that F_0 is the most salient voice femininity marker.[10–15] Research has also determined that there is an overlap between male and female speaking fundamental frequencies.[16] The lowest F_0 where listeners perceptually identified TWs as women in three separate studies was 156 Hz[17] and 165 Hz.[11,18] This range, with a slight variation, is known as gender ambiguous or gender neutral, and therapy for TW targets establishing greater than 170 Hz.[11] However, additional listening studies have revealed that even when F_0 is raised to a gender-ambiguous range, the voices of TW are often still perceived as male,[15,19,20] suggesting that listeners are aware of conflicting voice parameters. This finding has led to investigating other aspects of voice and speech that differ between men and women to employ as therapy targets for women who are TG. The most common parameters are voice quality (ie, signal complexity), pitch range and inflections (ie, pitch sigma), and resonance (ie, vowel FFs), of which the latter has garnered the most attention.[2,10,12,14,15,20–26]

Investigating the relationship between F_0 and FFs and gender perception has gained momentum since Peterson and Barney,[27] and later Hillenbrand et al,[16] published vowel FF mean values for groups of men and women and children. Some of the research has been theoretical in nature,[2,3,12,20,25,26,28,29] whereas other research has had a clinical focus for speech and voice therapy for individuals who are TG.[6,8,10,11] Table 1 provides an overview

Accepted for publication June 27, 2017.

From the *Speech-Language-Hearing Sciences, Loyola University MD, Baltimore, Maryland; †Language and Voice Experience, Rockville, Maryland; and the ‡Whisker Labs, Germantown, Maryland.

Address correspondence and reprint requests to Sally J.K. Gallena, Department of Speech-Language Pathology, Loyola University MD, 8890 McGaw Rd., Columbia, MD 21045. E-mail: sgallena@loyola.edu

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

http://dx.doi.org/10.1016/j.jvoice.2017.06.023

TABLE 1.
Summary of Pertinent Gender Perception Listening Study Methods, Stimuli, and Findings

Study	Original Source and Filter	Manipulations to Original Samples	Final Listening Stimuli	Findings
Schwartz and Rhine (1968)	No source-human filter	No manipulations	Whispered vowels /i/ and /ɑ/ Male and female speakers	FFs most salient cue for gender perception
Coleman (1976)	Synthetic source-human filter Electrolarynx as source	Source shifted	5 s of running speech Male and female speakers F_0 adjusted to two settings: 120 and 240 Hz	F_0 and FF cues for maleness were stronger than cues for femaleness for gender perception.
Whiteside (1998)	Synthesized source-synthesized filter Synthesized using Sensyn	Mismatched fundamental and FE	10 isolated vowels from running speech (50- or 100-ms duration) Male and female speakers Four conditions: F_0 and FF male F_0 and FF female F_0 male and FF female F_0 female and FF male	F_0 was the most salient cue for gender perception for 36 of 40 vowel samples.
Gelfer and Mikos (2005)	Synthesized source-human filter Synthesized using Dr. Speech	Mismatched fundamental and FEs	Isolated vowels: /i/, /ɜ˞/, and /u/ Male, female, and transgender speakers Two conditions: F_0 shifted to 120 and 240 Hz FFs unchanged	F_0 most salient cue for gender perception
Smith and Patterson (2005)	Synthesized source-synthesized filter Synthesized using STRAIGHT vocoder	Degrees of mismatch between source (GPR) and filter (VTL)	Isolated vowels (/ɑ/, /e/, /i/, /o/, and /u/) One male speaker Synthesized conditions: GPR shifted in six logarithmic steps from 61 to 523 Hz VTLs indicative of adult humans 2–12 ft tall	GPR and VTL both salient cues for gender perception when in "typical" ranges and configurations In mismatched configurations, VTL becomes a more salient cue than GPR.
Assmann et al (2006)	Synthesized source-synthesized filter Synthesized using STRAIGHT vocoder	Degrees of matched and mismatched fundamental and FF geometric means	Two sentences Male and female speakers Scaled in 10 steps for F_0 range of 60–450 Hz Mean $FF_{1–3}$ range of 850–2500 Hz	Accurate gender perception when F_0 and FF were typical of men and women True gender still perceived, suggesting other factors influence gender perception

(*continued on next page*)