JSLHR

Research Article

Perceptual Evaluation of Speech Naturalness in Speakers of Varying Gender Identities

Brandon Merritt[a] and Tessa Bent[a]

Purpose: The purpose of this study was to investigate how speech naturalness relates to masculinity–femininity and gender identification (accuracy and reaction time) for cisgender male and female speakers as well as transmasculine and transfeminine speakers.

Method: Stimuli included spontaneous speech samples from 20 speakers who are transgender (10 transmasculine and 10 transfeminine) and 20 speakers who are cisgender (10 male and 10 female). Fifty-two listeners completed three tasks: a two-alternative forced-choice gender identification task, a speech naturalness rating task, and a masculinity/femininity rating task.

Results: Transfeminine and transmasculine speakers were rated as significantly less natural sounding than cisgender speakers. Speakers rated as less natural took longer to identify and were identified less accurately in the gender identification task; furthermore, they were rated as less prototypically masculine/feminine.

Conclusions: Perceptual speech naturalness for both transfeminine and transmasculine speakers is strongly associated with gender cues in spontaneous speech. Training to align a speaker's voice with their gender identity may concurrently improve perceptual speech naturalness.

Supplemental Material: https://doi.org/10.23641/asha.12543158

Speech naturalness has been conceptualized as a listener's scaling between an individual's speech and the listener's representation of typical speech patterns, including the dimensions of rate, rhythm, intonation, stress patterning, and syntactic structure (Anand & Stepp, 2015; Ingham & Packman, 1978). The reliability of listeners' speech naturalness ratings has been investigated in individuals with fluency or motor speech disorders and in those utilizing tracheoesophageal speech (Anand & Stepp, 2015; Eadie & Doyle, 2002; Klopfenstein, 2015; Metz et al., 1990; Southwood, 1996). Within these disordered populations, listeners' ratings of more versus less natural speech align with typical versus atypical speech patterns (Anand & Stepp, 2015; Eadie & Doyle, 2002; Klopfenstein, 2015; Metz et al., 1990; Southwood, 1996). Even within typical speakers (e.g., children of various ages or adults from different dialect regions), listeners are sensitive to and can reliably rate speech naturalness (Coughlin-Woods et al., 2005; Mackey et al., 1997).

Relationships have been observed between listeners' naturalness ratings and a number of acoustic and auditory–perceptual measures including fundamental frequency (f_o), intensity range, voice onset time, articulation rate, syllable duration, strength of speaker dialect and foreign accent in relation to Standard American English, speaker intelligibility, and speech disfluencies (Klopfenstein, 2015; Mackey et al., 1997; Metz et al., 1990; Southwood, 1996). For example, more perceptually natural speech has been associated with an increased speaking rate for speakers with dysarthria (Klopfenstein, 2015) but a decreased speaking rate for speakers who stutter (Metz et al., 1990). Furthermore, Klopfenstein (2015) found mean f_o and syllable duration to both negatively and positively correlate with speech naturalness for speakers with dysarthria. That is, there was variability in the direction of the effects among speakers of similar speech characteristics. Therefore, speech parameters that contribute to the perception of naturalness appear to vary both across and within speaker populations.

It is also likely that multiple speech parameters interact to give rise to the perception of more versus less natural-sounding speech. For instance, Klopfenstein's (2015) finding of both positive and negative correlations between mean f_o and naturalness ratings for speakers with dysarthria may have been a consequence of the interaction with other parameters beyond mean f_o, such as f_o range, vocal quality, or rate of speech. Furthermore, although the

[a]Department of Speech, Language and Hearing Sciences, Indiana University Bloomington

Correspondence to Brandon Merritt: bmmerrit@iu.edu

Editor-in-Chief: Bharath Chandrasekaran
Editor: Amanda I. Gillespie

Received November 13, 2019
Revision received February 6, 2020
Accepted March 23, 2020
https://doi.org/10.1044/2020_JSLHR-19-00337

Disclosure: The authors have declared that no competing interests existed at the time of publication.

acoustic–phonetic variables included across studies to explain perceived speech naturalness have accounted for up to 71% of the variability in naturalness ratings (Hardy et al., 2016, 2020; Metz et al., 1990; Southwood, 1996), it remains possible that other unexplored variables may be contributing to the perception of naturalness in speech.

Speech Naturalness in Transgender Speakers

Speech naturalness has been an area of recent interest in transgender communication research following calls for this construct to be addressed as part of a comprehensive management plan for the transgender and gender nonconforming voice (Adler et al., 2012; Davies & Goldberg, 2006; Gelfer, 1999; Gelfer & Bennett, 2013).[1] A large proportion of individuals who identify as transgender report voice and language to be as important as gender affirmation surgery in achieving congruency between their gender identity and gender presentation to others (van Borsel et al., 2000). The goal of speech and voice training for these individuals is to achieve both sufficiently masculine/feminine and natural speech patterns for authentic voice presentation (Davies & Goldberg, 2006). Transgender individuals themselves report that natural speech and voice is a crucial component of gender presentation (Byrne, 2007). Hence, the perceptual implications of speech naturalness are of particular importance for transgender speakers: Speech that does not sound natural draws undesirable attention and may potentially "out" the speaker. Developing gender-affirming, natural speech may assist transgender individuals to increase their gender alignment, which is known to lead to greater affirmation of gender identity and more fluid social interactions (Davis & Meier, 2014).

Recent studies have begun to explore the acoustic parameters that contribute to natural speech for transfeminine speakers. In Hardy et al.'s (2016) study, greater perceived naturalness for transfeminine speakers was correlated with lower minimum f_o achieved during a frequency range task, higher second formant of the vowel /a/, and lower vocal shimmer percentage. Collectively, however, these variables only accounted for 44% of the variability in naturalness ratings for the transfeminine speakers. In a follow-up study, lower mean f_o, higher average formant values, and faster rate of speech were found to be predictive of transfeminine speaker naturalness (Hardy et al., 2020). Similarly, only 36% of the variance in speech naturalness ratings was explained by the three variables. Hence, other acoustic–phonetic features are likely needed to explain the variability in naturalness ratings for transgender speakers.

Although Hardy et al.'s work has provided substantial insight into the acoustic–phonetic factors that influence perceived naturalness for transgender speakers, the stimulus materials have been limited to controlled laboratory speech, including read speech. Compared to read speech, spontaneous speech is characterized by more variation in f_o contours, an increased speech rate, hypoarticulation, hesitations, repetitions, partial words, and disfluencies (Lieberman et al., 1985; Nakamura et al., 2008). Listeners are sensitive to these differences and able to classify speech as spontaneous versus read above-chance levels (Remez et al., 1985). As speech naturalness has been associated with variation of f_o, intensity range, speech rate, and disfluencies (Hardy et al., 2020; Klopfenstein, 2015; Mackey et al., 1997; Metz et al., 1990), differences between the acoustics of spontaneous and read speech may impact listener perception of naturalness. For instance, average f_o during an expository speech task was found to be a significant predictor of naturalness for transgender speakers in Hardy et al. (2020) but not found to be a significant predictor during a reading speech task in Hardy et al. (2016). The differences in results across these studies may be related to task differences. Additionally, because spontaneous speech is where most interactions occur between communication partners, spontaneous speech samples provide maximum ecological validity for the assessment of speech naturalness.

Evaluations of speech naturalness in individuals who identify as transgender have also been constrained to the evaluation of transfeminine speakers and have not included transmasculine speakers. Transmasculine individuals constitute a distinct speaker group from transfeminine speakers, in part, because they often are undergoing exogenous testosterone therapy as part of the gender affirmation process. Supraphysiologic doses of testosterone are known to elicit morphological changes of the vocal folds (Talaat et al., 1987), which consequently lower f_o (Irwig et al., 2017; Nygren et al., 2016). Because the gender identity of transmasculine speakers tends to be identified more accurately than transfeminine speakers (Azul et al., 2018; Scheidt et al., 2004), transmasculine speakers may be rated as having more natural-sounding speech. Additionally, the potential source–filter interaction (Fant, 1960) of the vocal folds, which are morphologically masculinized, with an upper airway of female-typical anatomy may have differential effects on perceived naturalness for transmasculine speakers as compared to transfeminine speakers, whose vocal folds and vocal tract remain male-typical. Thus, evaluation of spontaneous speech samples and speech by transmasculine speakers would provide a more comprehensive picture of perceptual naturalness in the transgender speaker population.

Perception of Gender and Masculinity/Femininity

Gender is a salient speaker attribute conveyed to listeners in the speech signal. Adult listeners can accurately

[1] We use the term "transgender" to refer to individuals who have a gender identity or gender expression that differs from their assigned sex at birth (Schilt & Westbrook, 2009). This term is in contrast to "cisgender," which refers to people whose gender identity matches the sex they were assigned at birth. Individuals assigned male at birth who identify along the female spectrum are described as "transfeminine," while individuals assigned female at birth who identify along the male spectrum are described as "transmasculine" (Azul et al., 2017). We use the term "gender identity" to refer to the gender that individuals wish to convey to others (e.g., male or female) and "gender orientation" to describe the match or mismatch of individuals' gender identity and the sex they were assigned at birth (i.e., cisgender or transgender).

identify cisgender speakers' gender from limited acoustic–phonetic information, including isolated voiceless fricatives, whispered vowels, low-pass filtered speech, and stimuli of less than two phonatory cycles (Lass et al., 1980; Lass et al., 1976; Schwartz, 1968; Schwartz & Rine, 1968). The significance of speaker gender in speech perception is further exemplified by findings that humans categorize voices as male or female from infancy (Miller et al., 1982). Even fetuses can discriminate between highly contrastive male and female voices in utero (Lecanuet et al., 1993).

In addition to gender identification accuracy, reaction time data from two-alternative forced-choice (2AFC) gender identification tasks have revealed significant listener response differences based on gender prototypicality of cisgender speakers (Babel & McGuire, 2015; Skuk & Schweinberger, 2014; Strand, 2000). Voices that are most prototypically male or female are categorized most quickly; voices that diverge from listener gender expectations appear to increase processing load and result in slower responses (Strand, 2000). Although no study has included a reaction time measure for gender identification with transgender voices, the evidence from studies of cisgender speakers suggests that the voices of speakers who diverge from prototypical male/female speech may result in additional cognitive load. This additional processing load may have additional perceptual consequences, such as lower perceived speech naturalness.

Given the salience of speaker gender from voice/speech to listeners and the constraints imposed by vocal tract anatomy, transgender speakers often face considerable challenges in achieving speech that is quickly and accurately identified as their gender identity. Transgender speakers are often misgendered and rated as less masculine or feminine sounding than corresponding cisgender speakers (Azul et al., 2018; Gelfer & Schofield, 2000; Gelfer & Tice, 2013; Gelfer & Van Dong, 2013; Hancock et al., 2014; Hancock & Pool, 2017; King et al., 2012; Schwarz et al., 2018). For example, across several studies, transfeminine speakers have been identified as their gender identity only between 20% and 30% of the time (Gelfer & Schofield, 2000; Hancock et al., 2014; Hardy et al., 2020; King et al., 2012). When evaluating transfeminine speakers who have had no history of communication feminization training, identification accuracy is even lower (Gelfer & Tice, 2013; Hardy et al., 2016). The gender identity of transmasculine speakers is identified more accurately than that of transfeminine speakers, partially due to the voice virilization associated with testosterone supplementation (Azul et al., 2018; Scheidt et al., 2004). They, however, still receive correct gender identification less than 80% of the time, an accuracy rate that is much lower than the 95%–100% gender identification accuracy typically seen for cisgender speakers (Hancock et al., 2014; Schwartz & Rine, 1968). Furthermore, ratings on a masculinity/femininity scale for transmasculine speakers have been found to be in a similar range as ratings for transfeminine speakers (Hancock et al., 2014).

For both cisgender and transgender speakers, f_o appears to be the strongest acoustic–phonetic characteristic associated with the perception of gender and masculinity/femininity (Gelfer & Mikos, 2005; Gelfer & Schofield, 2000; Hardy et al., 2020; Houle & Levi, 2019; Lass et al., 1976; Leung et al., 2018; Skuk & Schweinberger, 2014). However, vocal tract resonance (i.e., the filter function of the vocal tract applied to the sound wave produced by vocal fold vibration) also contributes important cues to gender. In particular, the first three vowel formants provide salient information regarding speaker gender and masculinity/femininity (Carew et al., 2007; Gallena et al., 2018; Gelfer & Bennett, 2013; Houle & Levi, 2019; King et al., 2012; Leung et al., 2018; Mount & Salmon, 1988; Skuk & Schweinberger, 2014), with males typically having lower and more closely spaced vowel formant frequencies compared to females (Kent & Vorperian, 2018; Peterson & Barney, 1952). Altering formant frequencies while maintaining f_o changes listener perception of speaker gender and perceived masculinity (Gallena et al., 2018; Kawitzky & McAllister, 2020). Listeners rely particularly heavily on formant information when f_o is in a gender-ambiguous range (i.e., 145–165 Hz; Gelfer & Bennett, 2013). Additional acoustic features such as intonation and vocal intensity (Hancock et al., 2014; Holmberg et al., 2010; Wolfe & Ratusnik, 1990) are also related to the perception of gender and masculinity. However, their influence appears secondary to f_o and formant frequencies, which together provide the most salient acoustic cues to speaker gender and perceived masculinity/femininity (Leung et al., 2018; Oates & Dacakis, 2015).

The acoustic–phonetic cues found to be important for the identification of speaker gender and perceived masculinity/femininity show a high degree of overlap with the cues for perception of speech naturalness. The association of these parameters with both gender cues (Gallena et al., 2018; Gelfer & Bennett, 2013; Gelfer & Mikos, 2005; King et al., 2012; Lass et al., 1976; Mount & Salmon, 1988; Skuk & Schweinberger, 2014) and perceived naturalness (Hardy et al., 2020) suggests that there may be interactions between perceived speech naturalness and gender cues, particularly for speakers who present with nonprototypically masculine/feminine voices. Here, we investigate how perceived naturalness relates to gender identification accuracy and ratings of masculinity/femininity.

Scales of Measurement: Metathetic Versus Prothetic

The study of human sensation and perception divides constructs into qualitative (metathetic, a substitutive quality) or quantitative (prothetic, an additive quality) perceptual continua (Stevens, 1975). To determine a construct's psychometric nature, ratings from two types of scales—equal-appearing interval (EAI) and direct magnitude estimation (DME)—are compared. In an EAI rating task, listeners rate stimuli with an equally spaced scale (e.g., 7 or 9 points). In a DME rating task, listeners provide a number value that they deem appropriate to represent a ratio of each stimulus to a standard. For a metathetic continuum interval scale, the ratings from the two tasks are linearly related, and either an EAI or DME scale would be appropriate for their measurement. In contrast, when listeners attempt to partition a prothetic construct into equal intervals, there is a systematic bias toward separating the lower end of the continuum into smaller intervals than the higher end (Stevens, 1975). This bias results in a curving of the trend line at higher values. A DME scale avoids the systematic bias associated with EAI scaling and is appropriate for use in measuring both metathetic and prothetic constructs (Stevens, 1975).

The comparison of EAI and DME scales to determine the rating task best suited to measuring perceptual speech naturalness has yielded inconsistent results. Naturalness has been found to be metathetic for tracheoesophageal speakers and individuals who stutter (Eadie & Doyle, 2002; Metz et al., 1990) but prothetic for speakers with dysarthria related to amyotrophic lateral sclerosis (Southwood, 1996). These findings suggest that listeners may rely on differing acoustic–phonetic cues for judgments of speaker naturalness across populations. These cues may vary in their psychometric nature and thus influence which rating task can most reliably measure naturalness within specific populations. Comparison of EAI versus DME scales has not yet been applied to speech naturalness from transgender speakers.

Research Questions

The primary objective of this project was to assess how perceived speech naturalness relates to ratings of masculinity/femininity and identification of speaker gender (accuracy and reaction time) for cisgender and transgender speakers using spontaneous speech. To this end, we addressed the following questions:

1. Which rating scale is most appropriate for measuring speech naturalness in speakers of varying gender identities?
2. Are there significant differences in perceived speech naturalness, perceived masculinity/femininity, and gender identification based on speaker gender identity and gender orientation?
3. How does speech naturalness relate to gender identification accuracy and reaction time?
4. How does speech naturalness relate to perceived masculinity/femininity?

Method

Stimuli

The stimuli included 160 utterances from 40 different speakers selected from podcasts available online by the first author. Podcast samples were chosen as stimuli because they allowed us to sample from a broad range of speakers and provided authentic, ecologically valid stimuli representative of speech most commonly encountered by listeners (for other research using podcast stimuli, see Davidson, 2019, and Lotfian & Busso, 2017). The speakers included 10 transfeminine speakers, 10 transmasculine speakers, 10 cisgender female speakers, and 10 cisgender male speakers. All speakers appeared to be native speakers of American English from the United States (Northeast = 14, West = 12, Southeast = 6, Southwest = 4, Midwest = 3, unknown region = 1) based on biographical information obtained online, when available. When this information was not available, the second author, who has nearly 20 years of experience in studying nonnative speech, listened to the samples and did not detect a noticeable nonnative accent for any speaker. Region of origin, race, age, and sexual orientation were determined by speakers' biographical information obtained online. Podcasts samples were first selected for transgender speakers and were identified by searching popular podcast platforms such as National Public Radio's Fresh Air and the Moth Radio Hour as well as aggregate podcast websites such as player.fm and stitcher.com. Each transfeminine and transmasculine speaker was matched as closely as possible on regional origin, race, and age with a cisgender female or male speaker, respectively. Once transgender speakers were selected, corresponding cisgender speakers were identified by searching the same databases for individuals who best matched the demographic profiles of the transgender speakers. Care was taken to select transgender and cisgender speakers who were not widely known. Speakers were White (n = 35), Asian American (n = 3), and African American (n = 2). The average age of speakers was 35 years (cisgender female: M = 41, range = 22–71; cisgender male: M = 33, range = 22–55; transfeminine: M = 38, range = 22–60; transmasculine: M = 32, range = 20–50). There were no significant age differences between speaker groups as determined by one-way analysis of variance (ANOVA), $F(3, 36) = 1.477$, $p = .23$. History of voice/communication training or history of hormone replacement therapy could not consistently be determined for speakers. Cisgender speakers did not identify as lesbian, gay, or bisexual. We attempted to control for sexual orientation due to evidence that a speaker's sexual orientation can influence their speech production patterns (Munson, 2007; van Borsel et al., 2013). Each speaker contributed four utterances that averaged 6.1 s (range: 3.2–8.5 s). Utterances were selected to start and end at the edges of prosodic units. Prosodic units were identified by pitch contour and breath groups in connected speech (Couper-Kuhlen, 2015). Utterances that included information that may have biased the listener (e.g., discussion of gender identity or transgender care or pertaining to personal relationships that may indicate gender identity or sexual orientation) and that was of perceptually or digitally poor audio quality (e.g., MP3 compression rate < 40 kbps; van Son, 2005), audible background speakers or other noises, muffled speech, audible distortion, utterances with excessive dysfluencies (e.g., hesitation, false starts, or stuttering), or utterances impersonating another person were excluded. Speakers were free of significant perceptual dysphonia as judged by the first author, a speech-language pathologist of over 10 years. Podcast episodes from which utterances were obtained were selected based on the presence of sufficient samples that met the above criteria. Each of the speakers' four utterances was sampled from the same podcast episode at 44.1 kHz. Praat (Boersma & Weenink, 2018) was used to equate utterances for root-mean-square amplitude and obtain f_o, semitone range, and speaking rate (de Jong & Wempe, 2009) for each utterance. Each speaker's values for these parameters were averaged across their four utterances. Ranges of these values for the four speaker groups were calculated based on the speaker with the highest average value and the speaker with the lowest average value within each respective group. This information is reported in Table 1.

To ensure that the utterances did not contain language (i.e., words or phrases) that would strongly cue perception of gender, utterances were evaluated by 10 adult monolingual American English speakers. The evaluators rated the written version of each utterance on whether they thought the speaker was male or female using a 7-point Likert scale where 1 = *definitely a male*, 7 = *definitely a female*, and 4 = *neutral*. They were instructed to use the middle range for language that was equally likely from a male or female speaker (i.e., "gender neutral"). All included stimuli received average ratings between 3 and 5, suggesting that raters did not perceive the written utterances as strongly gendered. Of the initial set of utterances selected, five did not meet this criterion. Therefore, an additional five utterances were collected from the same podcast episodes as the original utterances and subjected to the above rating task. All of these replacement utterances met the criterion. The utterance transcripts, their duration, gender ratings for the written utterances, and media source can be found in Supplemental Materials S1 and S2.

Listeners

Fifty-two monolingual American English listeners were recruited from the Indiana University campus and surrounding Bloomington community. Listeners (42 female, nine male, and one nonbinary) had an average age of 21.4 years (range: 18–32). Prior to the start of the experiment, listeners completed a language background and demographic questionnaire as well as a hearing screening. All listeners passed the hearing screening at 25 dB at 250 Hz and 20 dB at octave intervals between 500 and 8000 Hz (American National Standards Institute, 2010). No participant had taken more than two courses in communication sciences and disorders, and none had completed coursework pertaining to speech science or voice disorders. Listeners were not told that they would be listening to transgender speakers.

Rating Procedure

The research took place at the Indiana University Department of Speech, Language and Hearing Sciences after receiving approval from the institutional review board at Indiana University. Written consent was obtained from each listener. All listeners first completed a 2AFC gender identification task. A 2AFC task was selected over a scale that would capture a spectrum of gender identity to allow for comparison to prior work involving cisgender participants (Babel & McGuire, 2015; Skuk & Schweinberger, 2014). Following this task, listeners completed two rating tasks: one on masculinity/femininity and one on speech naturalness. Half of the listeners completed their ratings using a DME scale, and half used an EAI rating scale. The specific rating task (i.e., masculinity/femininity and naturalness) was counterbalanced across listeners. Stimulus presentation and response collection were controlled through PsychoPy 3.0.0 (Peirce et al., 2019) running on a Mac Mini with Sennheiser HD280 Pro headphones. Stimuli were presented at approximately 70 dB SPL. The experiment was conducted in a sound-attenuated booth with up to two listeners at a time. For each of the three tasks, participants were presented with all 160 stimuli once in a random order. Prior to the start of each task, listeners were presented with four practice trials. The stimuli for these practice trials included one utterance as produced by a speaker from each of the gender groups. Speakers in the practice trials were different than those included in the experiment. Listeners were given a 30-s break halfway through the gender identification and rating tasks, a 5-min break after completing the gender identification task, and a 3-min break between the naturalness and femininity rating tasks. Naturalness was not defined for listeners by the investigators in line with previous work (Metz et al., 1990); listeners were told to make their ratings based on what sounded more or less natural to them. The qualifying terms "very natural" and "very unnatural" were used in describing the naturalness rating scales to participants following the methods of Hardy

Table 1. Summary of acoustic measurements from the four speaker groups reported as means with ranges in parentheses.

Group	f_o, mean (range)	ST, mean (range)	Speaking rate (syll/s), mean (range)
Cisgender female	171 Hz (143–215)	21.46 (13.36–26.89)	4.91 (4.16–5.79)
Cisgender male	112 Hz (92–143)	13.51 (8.33–17.49)	5.01 (4.00–6.20)
Transfeminine	149 Hz (94–204)	15.59 (10.69–22.95)	5.10 (3.75–6.78)
Transmasculine	130 Hz (105–193)	14.92 (10.12–19.49)	4.84 (4.18–5.59)

Note. Each speaker's values were averaged across their four utterances. Ranges represent the speaker with the highest average value and the speaker with the lowest average value within each speaker group. Fundamental frequency is reported in hertz. Semitone (ST) mean and range were calculated using the formula: $39.86314 * \log_{10}(f_o\text{max} / f_o\text{min})$ (Henton, 1989). Speaking rate was calculated as the number of syllables divided by phonation time in seconds per utterance. syll/s = syllables per second.

et al. (2016, 2020) in their evaluation of naturalness for transgender speakers. After completing the naturalness rating task, listeners were asked to describe how they determined what sounded more or less natural. Participants were then debriefed as to the nature of the experiment. Because the experimental protocol required approximately 90 min for listeners to complete, stimuli were not repeated for reliability measurement.

Gender Identification

In the 2AFC gender identification task, listeners were instructed to categorize the speaker who produced each stimulus as male or female as quickly as possible. They indicated their responses on a Cedrus RB-740 response pad using their dominant hand to press one of two buttons labeled "male" and "female." Each trial began with a fixation cross, which was presented in the center of the screen for 500 ms, followed by 500 ms of a blank screen prior to the onset of the stimulus. After a participant entered their response, the next trial began. Reaction times were measured through PsychoPy from stimulus onset until a listener entered a response of "male" or "female" using the response pad. All listener reaction time values were at least 100 ms. The natural logarithm (Base 2.718) of reaction times was computed to normalize their distribution (Whelan, 2008).

Rating Task: EAI

For the EAI rating task, listeners were presented with a 9-point rating scale on each trial. Listeners selected a value to indicate their response. Once they selected a response, the number would appear on the screen. Then, the listener had to click on the number to confirm their response. For the naturalness rating block, 1 was labeled "very natural" and 9 was labeled "very unnatural." For the masculinity/femininity rating block, 1 was labeled "very masculine" and 9 was labeled "very feminine." Listeners were required to listen to the entire stimulus before they could enter a response. A 500-ms interstimulus interval with a blank screen was inserted between each stimulus.

Rating Task: DME—Without Modulus

For the DME rating task, listeners assigned a numeric value to each stimulus for the relevant dimension (i.e., perceived naturalness or femininity). DME was explained to listeners by showing them a picture of several parallel horizontal lines ranging in length from 0.32 to 25 cm and demonstrating how the length of the lines could be scaled with either DME or interval scaling (Stevens, 1975). Listeners practiced the DME procedure by scaling the length of several lines. Listeners were encouraged to scale the lengths of lines based on magnitudes of difference (e.g., "2, 4, 8, 16…" rather than "1, 2, 3, 4…"). It was explained that some lines were 2, 3, or 4 times as long as the smallest line and that number values they choose should reflect these magnitudes of difference. Listeners were then told that speech clips they hear may sound more or less natural or feminine than others based on such magnitudes of difference and that the number ratings they assign to speech clips should reflect these differences. Listeners were instructed to not use 0 as a rating. For the block in which listeners were rating femininity, they were instructed to use higher numbers for greater perceived femininity and lower numbers for less perceived femininity (i.e., greater perceived masculinity). For the block in which listeners were rating naturalness, listeners were instructed to use higher numbers for more unnatural-sounding speech. For both femininity and naturalness, DME without modulus was used; listeners assigned a value with any number they chose. They then scaled the femininity or naturalness of all subsequent speakers so that their ratings were proportional across speakers. An on-screen prompt asked listeners to provide a number value for how natural or feminine the speaker sounded to them after each speech clip finished playing. Listeners typed their number ratings for each stimulus using a keyboard. Listeners were required to listen to the entire stimulus before they could enter a response. A 500-ms interstimulus interval with a blank screen was inserted after each response entry.

Because the DME tasks did not use a modulus, DME values were modulus equalized to remove the error variance attributable to each listener's individual choice of modulus using the method in Lane et al. (1961). Total variance due to the listener's choice of modulus was removed by first computing the grand mean of all listener ratings. Next, each listener's mean for all their responses was subtracted from the grand mean. Finally, the resulting difference from this calculation was added back to each one of that listener's ratings. This procedure resulted in normally distributed interval-level data (Snow & Williges, 1998) with common modulus and was conducted for each of the 26 listeners.

DME values greater than 3 *SD*s from each listener's mean rating were removed to account for response entry errors (e.g., a rating of 67 when all other ratings were between 1 and 10 and averaged 4). All DME values of 0 were changed to 1 as geomean calculation cannot be completed with zero or negative numbers (Roenfeldt, 2018). Missing data points, ratings > 3 *SD*s above the mean, and 0-value DME ratings accounted for < 1% of responses. Pearson product–moment correlations were used to determine the relation between DME naturalness ratings and gender identification reaction times, gender identification accuracy, and DME femininity ratings across the 40 speakers, with each speaker's average score across their four utterances entered into the analyses.

Results

Rating Scale Comparisons

Following Stevens (1975), the arithmetic means of the EAI scale ratings were plotted against the geometric means of the DME scale ratings for each of the 40 speakers for both femininity and naturalness. Curvilinear regression was used to determine the model (quadratic or linear) that would best fit the data for both femininity and naturalness dimensions.

Naturalness

For naturalness ratings, the speakers' means from each group of 26 listeners who completed the EAI or DME rating tasks were calculated and compared. Based on a curvilinear regression analysis ($R^2 = .9169$), $F(2, 37) = 204.024$, $p < .05$), a second-order polynomial accounted for significantly more of the variance than the linear model (R^2 change = .012, line of best fit: $y = -0.0259x^2 + 0.795x + 1.2109$). Visual inspection of the plotted means revealed a slight downward-bowing curve of the regression line at higher values (see Figure 1). Thus, naturalness appears to be a construct that is best suited to measurement by a DME scale—subjective change in naturalness is not equal between ratings at lower and higher ends; there is bias toward dividing the lower end of the continuum into smaller intervals than the higher end. Therefore, DME naturalness ratings from the 26 participants who used this rating method were used for the analyses that follow.

Masculinity/Femininity

EAI arithmetic means for masculinity/femininity for each speaker were compared to their corresponding DME femininity geometric means (see Figure 2). Results indicated that there was a significant linear relationship between these two scales (line of best fit: $y = 0.4858x + 1.2699$, $R^2 = .967$), $F(1, 38) = 1{,}097.138$, $p < .001$. No significant improvement was found for the curvilinear model. Thus, either an EAI or DME scale would be appropriate to measure this construct. However, to maintain consistency across our naturalness and femininity rating scales, DME values were selected for use in the analyses below.

Between-Group Comparisons

Four repeated-measures ANOVAs were conducted to analyze (a) gender identification accuracy, (b) gender identification reaction times, (c) DME femininity ratings, and (d) DME naturalness ratings. Each ANOVA had two within-subject variables: gender identity (male vs. female) and gender orientation (cisgender vs. transgender).

Gender Identification Accuracy

Gender identification accuracy for the four speaker groups is shown in Figure 3. There were significant main effects of gender identity, $F(1, 51) = 244.1$, $p \le .001$, $\eta_p^2 = .827$, and gender orientation, $F(1, 51) = 4{,}035.1$, $p \le .001$, $\eta_p^2 = .988$. The main effect of gender identity arose because gender identification accuracy was higher for male-identified speakers than female-identified speakers. The main effect of gender orientation arose because cisgender speakers were correctly identified as their gender identity more often than transgender speakers. Additionally, there was a significant interaction between gender identity and gender orientation, $F(1, 51) = 250.2$, $p \le .001$, $\eta_p^2 = .831$. Listeners' performance for both cisgender male and female speakers was highly accurate with similar accuracy rates (99.8% and 97.8% correct, respectively). However, for the transgender speakers, listeners were less accurate overall with a difference across gender identities; listeners showed higher gender identification accuracy for speakers who were transmasculine (68.6% correct) than transfeminine (22.5% correct).

Gender Identification Reaction Time

Reaction times for the four speaker groups are shown in Figure 4. There was a significant main effect of gender orientation, $F(1, 51) = 203.2$, $p \le .001$, $\eta_p^2 = .799$. The main

Figure 1. Naturalness ratings for each of the 40 speakers with arithmetic means from the equal-appearing interval (EAI) scale on the *y*-axis and geometric means from the direct magnitude estimation (DME) scale on the *x*-axis.

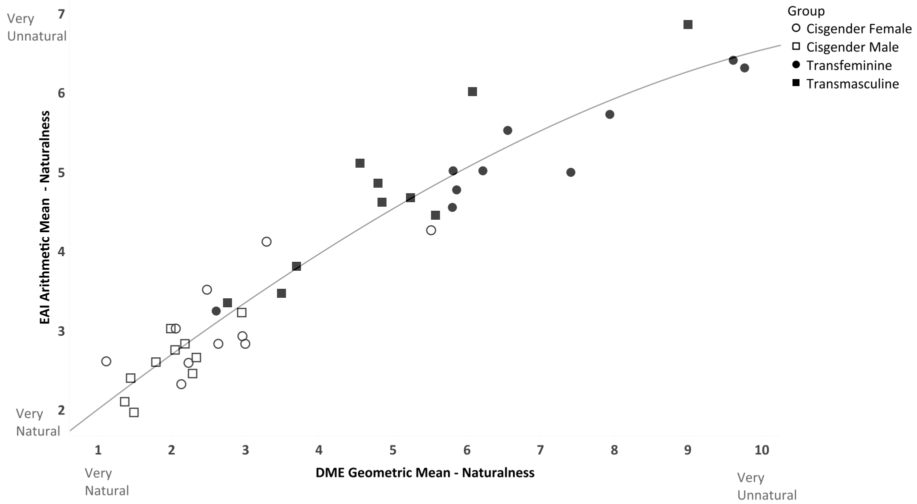
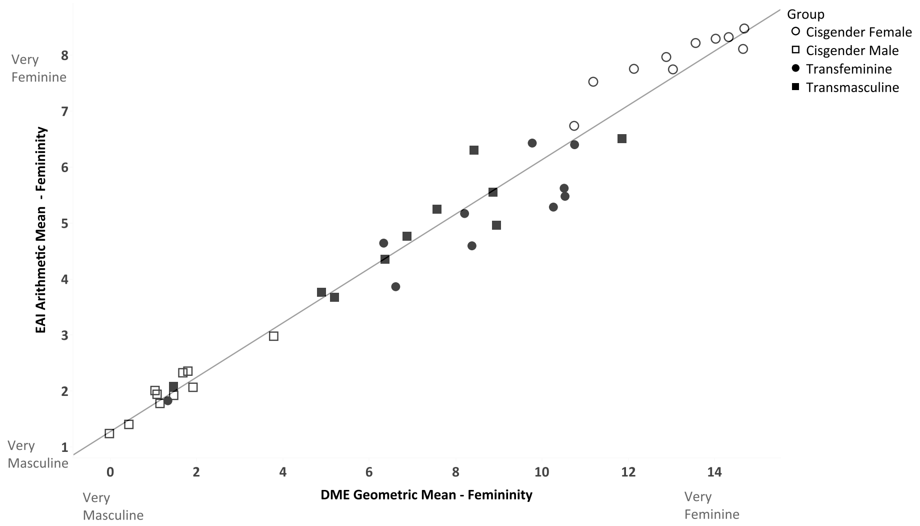


Figure 2. Arithmetic means of masculinity/femininity for the equal-appearing interval (EAI) scale are plotted on the *y*-axis, and geometric means of femininity ratings from the direct magnitude estimation (DME) scale are shown on the *x*-axis.



effect of gender orientation occurred because reaction times for transgender speakers were higher than those for cisgender speakers. There was not a significant effect of gender identity on listener reaction times. There was, however, a significant two-way interaction between gender identity and gender orientation, $F(1, 51) = 7.8$, $p \leq .01$, $\eta_p^2 = .132$. Listeners' reaction times for both cisgender male and female speakers were similar ($M = 624$, $SD = 351$, and $M = 634$, $SD = 336$ ms, respectively). However, for the transgender speakers, there was a difference across gender identities; listeners took longer to categorize transmasculine speakers as male or female ($M = 964$, $SD = 419$ ms) than transfeminine speakers ($M = 916$, $SD = 413$ ms).

DME Femininity Ratings

DME femininity ratings for the four speaker groups are shown in Figure 5. There was a significant main effect of gender identity on femininity ratings, $F(1, 25) = 16.9$, $p \leq .001$, $\eta_p^2 = .403$). Female-identified speakers were rated as more feminine than male-identified speakers. There was not a significant effect of gender orientation on femininity ratings. Again, there was a significant two-way interaction between gender identity and gender orientation on femininity ratings, $F(1, 25) = 8.4$, $p \leq .01$, $\eta_p^2 = .252$. Transmasculine and transfeminine speakers were rated similarly regarding femininity ($M = 7.0$, $SD = 4.7$, and $M = 8.2$, $SD = 6.5$, respectively), whereas cisgender males and females

Figure 3. Gender identification accuracy in proportion correct for the four speaker groups. The top and bottom of the boxes indicate the 75th and 25th percentiles, respectively. The solid line within the box marks the median. Error bars indicate the 90th and 10th percentiles.

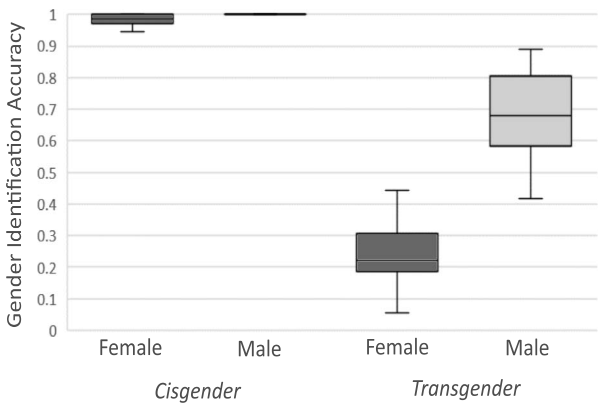


Figure 4. Listener reaction times to identify a speaker's gender as "male" or "female" for the four speaker groups. Reaction times are reported in log natural (ln) milliseconds.

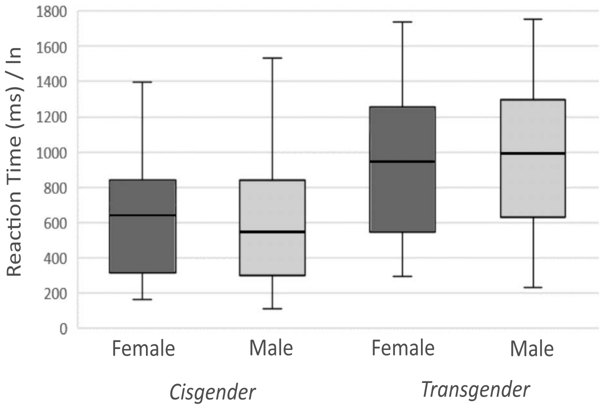
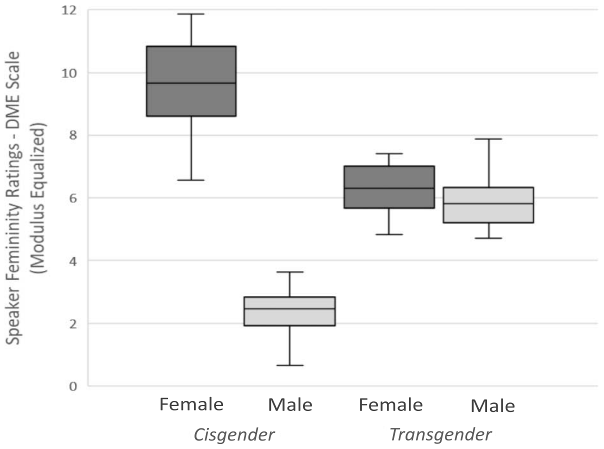


Figure 5. Direct magnitude estimation (DME) femininity ratings by speaker group. Values reported are modulus equalized. Higher values indicate higher rated femininity.

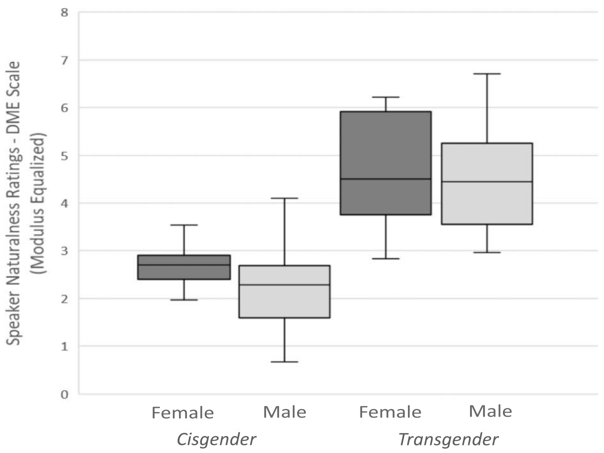


received substantially different femininity ratings (M = 1.4, SD = 3.0, and M = 13.1, SD = 13.7, respectively).

DME Naturalness Ratings

DME speech naturalness ratings for the four speaker groups are shown in Figure 6. There was a significant main effect of gender orientation on naturalness ratings, $F(1, 25) = 11.0$, $p \leq .01$, $\eta_p^2 = .305$. The effect of gender orientation arose because individuals who identified as transmasculine and transfeminine were rated significantly less natural sounding (M = 5.0, SD = 2.7, and M = 6.7, SD = 7.9, respectively) than those who identified as cisgender male or female (M = 2.0, SD = 1.1, and M = 2.7, SD = 0.9, respectively). The main effect of gender identity did not reach significance, $F(1, 25) = 3.8$, $p = .063$, $\eta_p^2 = .132$. The two-way interaction was not significant.

Figure 6. Direct magnitude estimation (DME) speech naturalness ratings by speaker group. Values are modulus equalized. Higher values indicate less natural speech.



Qualitative Responses

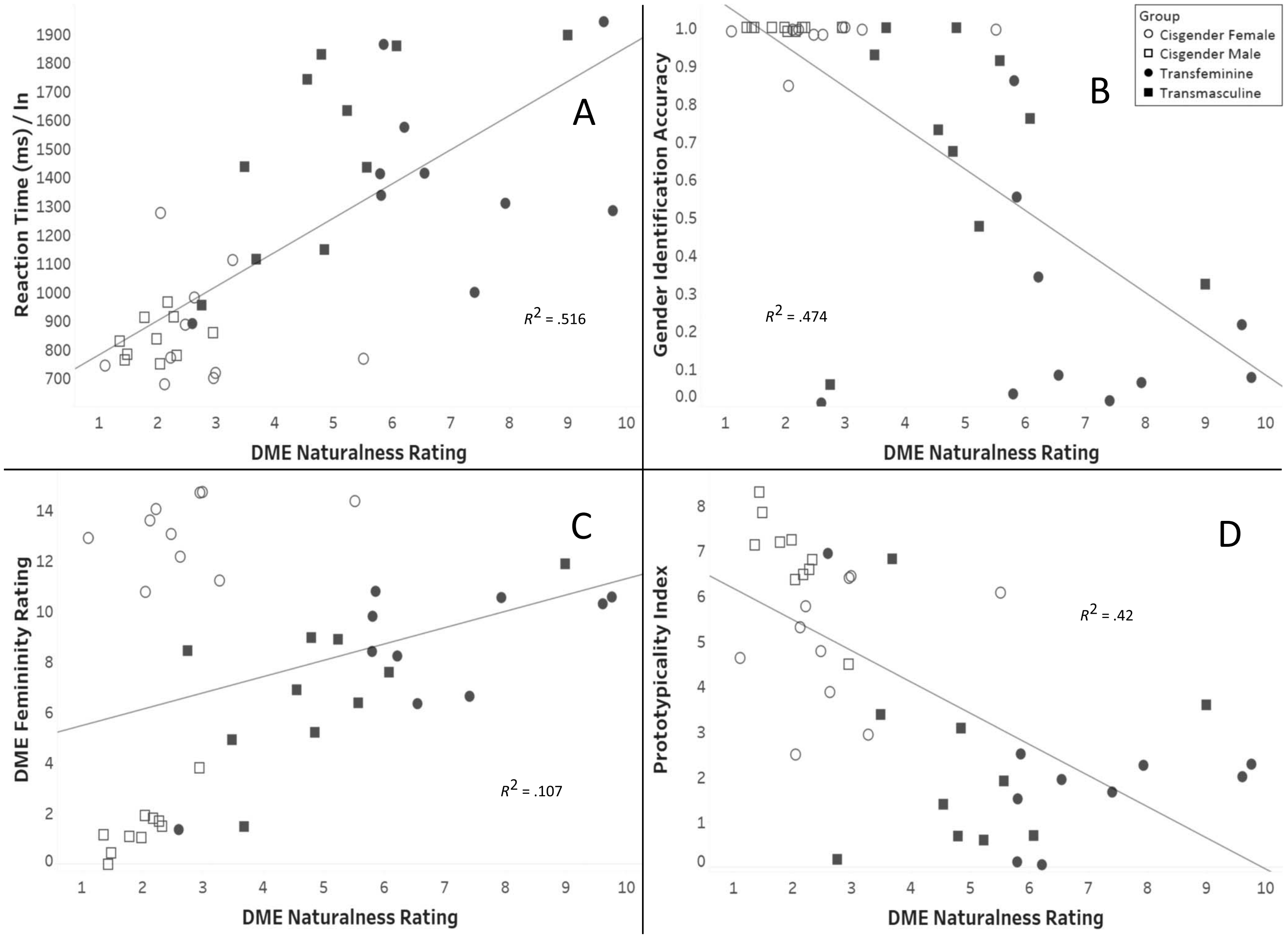
After completing the naturalness rating task, participants described what aspects of speech and/or voice they used to determine what sounded more or less natural. A thematic analysis (Braun & Clarke, 2012) was conducted by the first author to assess general themes. All listener responses were evaluated for key content words, and the percentage of responses that contained the same content themes was calculated. Themes included effort in changing speech (37%, e.g., "…if it sounded forced—like they were actively thinking about the way they were speaking or altering their way of talking."), rate/prosody (35%, e.g., "…if they used inflections in their voice, or the smoothness in their speech was natural…their fluency."), gendered voice/speech (29%, e.g., "voices that sounded more masculine were more natural"), regional dialect (25%, e.g., "how often I've heard their accent before"), vocal quality (24%, e.g., "breathy or creaky voices sounded more unnatural"), and nasalilty/resonance (10%, e.g., "less natural if they sounded more nasal").

Relationships Between Naturalness, Gender Identification, and Masculinity/Femininity

Pearson product–moment correlations between DME naturalness ratings and gender identification reaction times, gender identification accuracy, and DME femininity ratings were calculated (see Figure 7). Gender identification reaction times were strongly correlated with naturalness ratings, $r(38) = .718$, $p < .001$. Speakers who were more quickly categorized as male or female tended to be perceived as more natural sounding. Gender identification accuracy scores were also strongly negatively correlated with naturalness ratings, $r(38) = -.688$, $p < .001$. Speakers who were less accurately identified as their gender identity tended to be perceived as less natural sounding.

DME speech naturalness and DME femininity ratings showed a nonlinear, nonmonotonic relationship. This relationship arose because speakers whose ratings were at either extreme of the femininity rating scale (either very masculine or very feminine) tended to be perceived as more natural sounding. To address the nonlinear relationship, we transformed the masculinity/femininity data to create a new Prototypicality Index. The idea behind this index was to create a scale that equated speakers rated at either extreme end of the scale versus those rated more centrally and allow a linear correlation measurement between naturalness and masculinity/femininity to be calculated. To create this measure, the median value of DME ratings for all 40 speakers was first determined (8.3). Then, the absolute value of the difference between the median value and each speaker's mean DME femininity rating was determined, giving rise to their Prototypicality Index. Higher Prototypicality Index values for speakers indicate that listeners tended to rate them at the extremes of the DME rating scale—that is, very masculine or very feminine. Lower Prototypicality Index values for speakers indicate that listeners rated them closer to the median rating and therefore as more gender

Figure 7. Scatter plots of direct magnitude estimation (DME) naturalness ratings by (A) reaction time in log natural (ln) milliseconds, (B) gender identification accuracy, (C) DME femininity ratings, and (D) Prototypicality Index for each speaker group.



ambiguous. The mean Prototypicality Index value for each speaker group was 4.9 for cisgender female, 6.8 for cisgender male, 2.1 for transfeminine, and 2.2 for transmasculine. There was a strong negative correlation between naturalness ratings and the Prototypicality Index, $r(38) = -.648$, $p \leq .001$ (see Figure 7). Speakers who were perceived as less gender ambiguous tended to be rated as more natural sounding.

Discussion

This study investigated how perceived speech naturalness relates to masculinity–femininity ratings and speaker gender identification (accuracy and reaction time) for cisgender and transgender speakers. Because DME scaling resulted in a significantly better-fitting model over EAI scaling for perceived naturalness, DME scores were used for subsequent analyses. The DME responses indicated that transfeminine and transmasculine speakers were rated as significantly less natural sounding than cisgender female and male speakers. The gender identity of voices rated as less natural also were identified less accurately and more slowly. Finally, less natural-sounding voices were rated as less prototypically masculine/feminine. Listeners' qualitative responses for how they determined naturalness largely aligned with parameters associated with perceptual naturalness in stuttering, dialects differing from Standard American English, and tracheoesophageal speakers (e.g., rate/prosody, dialectal variation, vocal quality; Coughlin-Woods et al., 2005; Mackey et al., 1997; Metz et al., 1990) but also revealed criteria that have not, to our knowledge, been reported in previous evaluations of perceptual speech naturalness (e.g., effort in changing speech, gendered voice/speech, and nasality/resonance).

These findings suggest that there is a link between perception of speech naturalness and gender prototypicality. Specifically, naturalness ratings were strongly correlated with the perceptual ratings of masculinity/femininity as measured by our Prototypicality Index: Voices rated at the extreme ends of the masculine/feminine scale tended to be rated as more natural sounding than voices that received more intermediate ratings. The intermediate ratings on this scale suggest that these speakers were perceived as more gender ambiguous. Most of the transgender speakers received these intermediate ratings on the masculinity/femininity scale, suggesting that these speakers diverged from listeners' expectations of what prototypical male or female speech sounds like. This divergence from expected gender norms may have been one of the factors contributing to the overall lower naturalness ratings for the transgender speakers compared to the cisgender speakers. This interpretation is supported by listeners' descriptions of how they determined more versus less natural-sounding speech. Many listeners described speech that sounded more gender ambiguous or voices perceived as male with higher pitches as less natural sounding. Specifically, listeners reported that they rated voices that sounded "less masculine," "more feminine than they should," or "in the androgynous range" as less natural. These qualitative data support our quantitative findings that gender cues in speech have strong relations with perceived speech naturalness. The gender identification results also support the link between perception of gender cues and perceived naturalness: Speakers who were more frequently identified as their gender identity tended to be rated as more natural sounding. The lack of correct gender identification impacts transgender speakers much more often than cisgender speakers, almost all of whom had gender identification accuracy scores at ceiling in this study. In contrast, our rates of gender identification accuracy for transfeminine speakers (23%) were much lower and similar to previous studies, which have reported identification accuracy for transfeminine speakers between 2% and 30% (Gelfer & Schofield, 2000; Gelfer & Tice, 2013; Hancock et al., 2014; Hardy et al., 2016, 2020; King et al., 2012). Gender identification accuracy for transmasculine speakers in this study (69%) was higher than that for transfeminine speakers but still lower than that for cisgender speakers. These results align with previous findings for transmasculine speakers (50% correct in Scheidt et al., 2004, and 71% correct in Azul et al., 2018). Our findings thus confirm poor gender identification for transgender speakers and demonstrate the potential negative effect on speech naturalness.

Furthering the connection between perceived gender prototypicality and naturalness for transgender speakers is the significant correlation between gender identification reaction time and naturalness ratings. Longer reaction times for transgender speakers may have resulted from their divergence from prototypical male/female speech norms. Similar results have been reported in the cisgender literature, where listeners also took longer to categorize speakers who were rated as less prototypically male/female in speeded gender identification tasks (Babel & McGuire, 2015; Strand, 2000). In this study, divergences from gender prototypicality potentially resulted in increased listener cognitive load when evaluating transgender voices. Listeners may have interpreted an increase in cognitive effort to determine speaker gender identity as an indication of perceptually less natural speech when rating this construct.

Gender identification reaction times differed between transfeminine and transmasculine speakers. Listeners took significantly less time to categorize transfeminine speakers as "male" or "female" than transmasculine speakers. This discrepancy may have arisen because transfeminine speakers were overwhelmingly categorized as male in the gender identification task. Many of the transfeminine speakers appeared to fall unambiguously into the "male" category (i.e., six of the 10 speakers were identified as male > 90%), and therefore, listeners were fast, but highly inaccurate, when identifying the gender identity of these speakers. Because there were no visual cues to counter listeners' decisions, they may have been more confident (and consequently faster) in their categorization of transfeminine speakers as male when compared to their categorization of transmasculine speakers. Indeed, transfeminine speakers have been rated as more feminine when both auditory and visual cues were provided as compared to auditory-only cues (van Borsel et al., 2001). If listeners were provided with visual information that cued the speaker's gender identity (e.g., hairstyle, clothing), we anticipate that gender identification reaction times for these transfeminine speakers would be substantially slower, as listeners would potentially be less likely to place the speakers squarely in the "male" category. The provision of such cues along with the speech could also have a substantial impact on listener naturalness ratings, in cases where there is a mismatch between gender identification based on speech versus other cues to gender. The influence of visual gender cues on perceived speech naturalness should be evaluated in future work.

Investigations of how listeners initially develop concepts of gender prototypicality may provide further support for the relation between perceived naturalness and gender prototypicality. Exemplar (or hybrid) models provide a mechanism for how listeners build both representations of speaker-specific categories and broader indexical categories, including gender (Goldinger, 1998; Pierrehumbert, 2016). Cumulative exposure to typical male and female speech patterns, which are based on gender normativity and culturally determined, leads to the development of models for male and female speech characteristics. When listeners encounter a new speaker, they must determine which prior experiences they should draw from to map the incoming acoustic signal to phonetic categories (Kleinschmidt & Jaeger, 2015). For speakers who do not fit within a listener's established distributions for either male or female speakers, the determination of which generative model to draw from likely requires additional work and consequently may lower the perceived naturalness of these speakers.

Because listeners' prior experiences shape the way they interpret incoming speech signals, listeners with social networks that include larger numbers of speakers who fall outside of the gender binary may have developed different gender cue distributions and therefore may have different conceptions of speech naturalness. Listeners who have more frequently encountered individuals with less prototypically male/female speech characteristics (e.g., speakers who are transgender or nonbinary) may develop broader generative models, or models beyond "male" and "female," which could include a gender-ambiguous category. Broader models would contain a wider range of speech characteristics, which listeners associate with male and female speaker groups. Consequently, a speaker whose speech characteristics (e.g., f_o or formant frequencies) fall outside of normative ranges for cisgender speakers may be more likely to be perceived by listeners with more gender diversity in their social networks as prototypically male or female and, hence, more natural sounding. Support for this possibility can be seen in Hancock and Pool (2017), where transfeminine speakers were perceived as significantly more feminine by lesbian, gay, and bisexual as compared to heterosexual listeners. Lesbian, gay, bisexual, and transgender individuals are more likely to have gender-diverse social networks (Queen, 1998). These more diverse social networks may result in models of gender-based speech characteristics that differ from the models in the minds of listeners with less gender-diverse social networks. There is also evidence that social network composition influences perception of other indexical factors from speech. For instance, a listener more accurately identifies a speaker's ethnicity if the listener's social network is more inclusive of that ethnicity (Wong & Babel, 2017). In our study, the vast majority of our speakers and listeners were White, and we did not account for the ethnic composition of our listeners' social networks. Future work should thus consider how the intersections among race, ethnicity, and gender identity may influence perceptual judgments of gender identity and speech naturalness.

In addition to listeners' social networks, listeners' gender identity and orientation may influence their perception of gender cues in speech. Support for this possibility is evident in recent research with cisgender individuals where speaker intelligibility was shown to be influenced by listener gender identity (Yoho et al., 2019). Moreover, in gender identification tasks, cisgender males and females have been found to categorize voices of the opposite sex more accurately than voices of their own sex (Junger et al., 2013). Recent neuroimaging studies examining perception of speaker masculinity and femininity based on gender orientation have revealed that transmasculine and transfeminine listeners demonstrate neural activation patterns and behavioral response analogous to, yet distinct from, cisgender males and females, respectively (Junger et al., 2014; Smith et al., 2018). These results indicate that transgender individuals may constitute a distinct listener group whose perceptual experiences may significantly differ from those of cisgender individuals. All but one of the listeners in our study was a cisgender male or female with many more female than male listeners. Therefore, it is not possible for us to determine how listener gender identity and orientation influence perception of speech naturalness. Future research should include listeners with other gender identities and orientations as well as measures of the gender diversity in all listeners' social networks to determine how these factors shape perception of naturalness and gender cues in speech. Inclusion of reliability measures of listeners' judgments would also strengthen future experimental designs.

A final issue deserving of attention in future work is the determination of the specific acoustic–phonetic cues that listeners use to determine what makes speech sound more or less natural. In Hardy et al. (2020), f_o was the strongest predictor of transfeminine speaker naturalness ratings as well as gender identification and masculinity/femininity ratings. Naturalness and f_o values for their speakers, however, demonstrated a negative relationship, presumably because most of their speakers were identified as males by listeners, and higher f_o for perceived male speakers could have been perceptually less natural sounding. In line with these findings, the average f_o of our cisgender male and female speakers generally fell within previously reported f_o ranges (107–146 Hz for males and 143–275 Hz for females; Hollien & Shipp, 1972; Stoicheff, 1981), with the exception of our male speakers' ranges extending slightly below these norms (see Table 1). In contrast, our transfeminine speakers' f_o ranges were below cisgender female norms, and the majority were identified as males. Transmasculine speakers' f_o ranges fell between cisgender male and female norms, and they were more ambiguously categorized (e.g., < 70% correct gender identification accuracy and similar femininity ratings to transfeminine speakers). Listeners in our study also took longer to determine the gender identity of transmasculine than transfeminine speakers. Hence, listeners appeared to have more difficulty placing transmasculine speakers squarely in the male or female category, which may have been, at least in part, based on the f_o information they were provided. Given the significantly lower naturalness ratings that our transgender speakers received as compared to cisgender speakers, it may be that f_o is an especially important feature of perceptual speech naturalness for transgender speakers who are not readily recognized as their gender identity. It is yet to be determined, however, what features make speech more perceptually natural for transgender individuals whose speech elicits high gender identification accuracy. In particular, it would be helpful for future work to evaluate perceptual naturalness of transgender speakers whose f_o values fall within norms of cisgender males and females with corresponding gender identities. Such evaluation would provide a clearer picture of the additional acoustic–phonetic cues beyond f_o that listeners may use to judge speech naturalness in these individuals.

Formant frequencies are another likely contributor to perception of speech naturalness, as has been shown for transfeminine speakers (Hardy et al., 2016, 2020). The direction of the relationship between formant frequencies and perceptual naturalness for transfeminine speakers, however, conflicts with the direction of the relationship seen with f_o: Higher formant frequencies relate to perceptually more natural speech (Hardy et al., 2020). Why higher formant frequencies would relate to more natural speech while higher f_o would relate to less natural speech for transfeminine speakers is not entirely clear. Additionally, there has been no published literature investigating relations between vocal resonance and naturalness for transmasculine speakers specifically. Consequently, the interaction of f_o and formant frequencies as it relates to perceptual naturalness in transgender speakers remains an underexplored and intriguing issue.

A number of clinical implications arise from our findings. First, because we found speech naturalness to be strongly tied to perception of speaker gender identity and masculinity/femininity, training to align voice with gender identity may positively affect perceived speech naturalness. However, aligning a client's voice with their gender identity or desired gender presentation may not result in a prototypical voice according to Western, cisgender expectations. We support training goals that are client centered and not focused on perpetuating norms or stereotypes. Furthermore, because a large percentage of our listeners reported perceptually forced or effortful speech to be less natural sounding, listeners may be sensitive to the degree of perceived effort in speech production to achieve gender congruence. Speech that is perceived as forced may be related to variations in rate, intonation, tone, and stress, given that many listeners reported forced speech to affect speech rhythm and flow. Indeed, rate of speech and upward intonation shifts were evaluated by Hardy et al. (2020) as parameters related to speech naturalness. In their work, rate of speech served as a significant predictor of naturalness for transfeminine speakers. Intonation shifts, however, were excluded from the regression analysis due to low intrarater reliability of their measurement. Specifically examining prosodic measurements of intonation, tone, and stress may better elucidate the contribution of these parameters to perceptually natural speech for transgender speakers.

Our finding that transmasculine and transfeminine speakers received similar naturalness and masculinity/femininity ratings indicates that both groups may face obstacles to achieving natural, gender-congruent speech. Yet, transmasculine individuals make up only a small percentage of transgender clients who seek speech and voice training for gender congruence (Davies et al., 2015). These disproportional numbers may be based on the expectation that use of testosterone therapy alone will provide sufficient voice change for satisfactory gender alignment (Descloux et al., 2012; Gorton & Erickson-Schroth, 2017; Pettit, 2004). In fact, several studies demonstrate that testosterone therapy alone may not lead to satisfactory voice change for many transmasculine individuals (Azul et al., 2018; Azul & Neuschaefer-Rube, 2019; Azul et al., 2017; Cosyns et al., 2014; Hancock et al., 2017; Nygren et al., 2016). Moreover, many transmasculine individuals choose not to take exogenous testosterone (Gooren et al., 2015). Our findings highlight the need for increased speech and voice service awareness and availability for individuals who identify as transmasculine (Azul, 2015; Azul et al., 2018, 2017). Speech and voice training is a critical component of the gender affirmation process, which should be available as part of a comprehensive care plan for both transmasculine and transfeminine individuals (Adler et al., 2012; Oates & Dacakis, 2015). Future intervention studies should evaluate whether training that targets gender alignment specifically also has an impact on naive listeners' perception of speech naturalness.

Because we used stimuli from podcasts, we were not able to account for speakers' history of speech or voice training and the dosage or timing of testosterone therapy in our transmasculine speakers, which is known to affect f_o and perception of transmasculine speaker gender identity (Nygren et al., 2016; Scheidt et al., 2004). Future work should aim to account for these variables as they relate to naturalness in transgender speakers. However, the use of podcast stimuli allowed us to sample from a broader range of speakers than have typically been used. In previous studies, speakers were generally selected from clients seeking voice therapy (Hancock et al., 2011; Hardy et al., 2016, 2020; King et al., 2012; Schwarz et al., 2018). Furthermore, our stimuli included free choice of lexical content, variable rate, and the presence of fillers, repetitions, and disfluencies, all of which are characteristics of naturally produced speech. These spontaneous speech samples may better reflect speech encountered in the real world compared to speech produced in the laboratory (e.g., read speech or expository speech produced in response to a specific prompt). However, we could not control speaker recording conditions including recording environment and equipment. These uncontrolled variables may have introduced acoustic variability across speaker recordings (e.g., sampling rate, signal-to-noise ratio, proximity effect, distortion) that may have influenced listeners' ratings. We attempted to reduce these possible influences by instructing listeners to make their naturalness ratings based on the speakers' speech/voices rather than the audio signal quality. Controlling for recording conditions would also allow for an additional acoustic analysis beyond what we present here, such as accurate measurement of vocal intensity, long-term average spectra, or cepstral coefficients.

In conclusion, speech naturalness is a quantifiable and salient feature of speech in individuals of varying gender identities. Listeners are sensitive to gender-based cues from speakers that are strongly associated with perceptual naturalness. Targeting salient gender markers in speech and voice training may concurrently address speaker naturalness and yield greater gender affirmation and fluent social interactions for transgender clients.

Acknowledgments

We would like to thank Karly Jones for assistance with data collection. We are also grateful to the Department of Speech, Language, and Hearing Sciences at Indiana University for providing funding in support of this research.

References

Adler, R., Hirsch, S., & Mordaunt, M. (2012). *Voice and communication therapy for the transgender/transsexual client: A comprehensive clinical guide* (2nd ed.). Plural.

American National Standards Institute. (2010). *Specifications for audiometers (ANSI S3.6-2010)*.

Anand, S., & Stepp, C. E. (2015). Listener perception of monopitch, naturalness, and intelligibility for speakers with Parkinson's disease. *Journal of Speech, Language, and Hearing Research, 58*(4), 1134–1144. https://doi.org/10.1044/2015_JSLHR-S-14-0243

Azul, D. (2015). On the varied and complex factors affecting gender diverse people's vocal situations: Implications for clinical practice. *SIG 3 Perspectives on Voice and Voice Disorders, 25*(2), 75–86. https://doi.org/10.1044/vvd25.2.75

Azul, D., Arnold, A., & Neuschaefer-Ruber, C. (2018). Do transmasculine speakers present with gender-related voice problems? Insights from a participant-centered mixed-methods study. *Journal of Speech, Language, and Hearing Research, 61*(1), 25–39. https://doi.org/10.1044/2017_JSLHR-S-16-0410

Azul, D., & Neuschaefer-Rube, C. (2019). Voice function in gender-diverse people assigned female at birth: Results from a participant-centered mixed-methods study and implications for clinical practice. *Journal of Speech, Language, and Hearing Research, 62*(9), 3320–3338. https://doi.org/10.1044/2019_JSLHR-S-19-0063

Azul, D., Nygren, U., Södersten, M., & Neuschaefer-Rube, C. (2017). Transmasculine people's voice function: A review of the currently available evidence. *Journal of Voice, 31*(2), 261.e9–261.e23. https://doi.org/10.1016/j.jvoice.2016.05.005

Babel, M., & McGuire, G. (2015). Perceptual fluency and judgments of vocal aesthetics and stereotypicality. *Cognitive Science, 39*(4), 766–787. https://doi.org/10.1111/cogs.12179

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* [Computer software]. http://www.praat.org/

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 57–71). American Psychological Association. https://doi.org/10.1037/13620-004

Byrne, L. A. (2007). *My life as a woman: Placing communication within the social context of life for transsexual women* [Unpublished doctoral dissertation]. La Trobe University.

Carew, L., Dacakis, G., & Oates, J. (2007). The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals. *Journal of Voice, 21*(5), 591–603. https://doi.org/10.1016/j.jvoice.2006.05.005

Cosyns, M., Van Borsel, J., Wierckx, K., Dedecker, D., Van de Peer, F., Daelman, T., Laenen, S., & T'Sjoen, G. (2014). Voice in female-to-male transsexual persons after long-term androgen therapy. *Laryngoscope, 124*(6), 1409–1414. https://doi.org/10.1002/lary.24480

Coughlin-Woods, S., Lehman, M. E., & Cooke, P. A. (2005). Ratings of speech naturalness of children ages 8–16 years. *Perceptual and Motor Skills, 100*(2), 295–304. https://doi.org/10.2466/pms.100.2.295-304

Couper-Kuhlen, E. (2015). Intonation and discourse. In D. Tannen, H. Hamilton, & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed., pp. 82–104). Blackwell. https://doi.org/10.1002/9781118584194.ch4

de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods, 41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Davidson, L. (2019). The effects of pitch, gender, and prosodic context on the identification of creaky voice. *Phonetica, 76*(4), 235–262. https://doi.org/10.1159/000490948

Davies, S., & Goldberg, J. M. (2006). Clinical aspects of transgender speech feminization and masculinization. *International Journal of Transgenderism, 9*(3–4), 167–196. https://doi.org/10.1300/J485v09n03_08

Davies, S., Papp, V. G., & Antoni, C. (2015). Voice and communication change for gender nonconforming individuals: Giving voice to the person inside. *International Journal of Transgenderism, 16*(3), 117–159. https://doi.org/10.1080/15532739.2015.1075931

Davis, S. A., & Meier, S. C. (2014). Effects of testosterone treatment and chest reconstruction surgery on mental health and sexuality in female-to-male transgender people. *International Journal of Sexual Health, 26*(2), 113–128. https://doi.org/10.1080/19317611.2013.833152

Descloux, P., Isoard-Nectoux, S., Matoso, B., Matthieu-Bourdeau, L., Schneider, F., & Schweizer, V. (2012). Transsexuality: Speech therapy supporting the "voice" of transformation. *Revue de Laryngologie-Otologie-Rhinologie, 133*(1), 41–44.

Eadie, T. L., & Doyle, P. C. (2002). Direct magnitude estimation and interval scaling of naturalness and severity in tracheoesophageal (TE) speakers. *Journal of Speech, Language, and Hearing Research, 45*(6), 1088–1096. https://doi.org/10.1044/1092-4388(2002/087)

Fant, G. (Ed.). (1960). *Acoustic theory of speech production*. Mouton.

Gallena, S. J., Stickels, B., & Stickels, E. (2018). Gender perception after raising vowel fundamental and formant frequencies: Considerations for oral resonance research. *Journal of Voice, 32*(5), 592–601. https://doi.org/10.1016/j.jvoice.2017.06.023

Gelfer, M. P. (1999). Voice treatment for the male-to-female transgendered client. *American Journal of Speech-Language Pathology, 8*(3), 201–208. https://doi.org/10.1044/1058-0360.0803.201

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

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice, 19*(4), 544–554. https://doi.org/10.1016/j.jvoice.2004.10.006

Gelfer, M. P., & Schofield, K. J. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *Journal of Voice, 14*(1), 22–33. https://doi.org/10.1016/S0892-1997(00)80092-2

Gelfer, M. P., & Tice, R. M. (2013). Perceptual and acoustic outcomes of voice therapy for male-to-female transgender individuals immediately after therapy and 15 months later. *Journal of Voice, 27*(3), 335–347. https://doi.org/10.1016/j.jvoice.2012.07.009

Gelfer, M. P., & Van Dong, B. R. (2013). A preliminary study on the use of vocal function exercises to improve voice in male-to-female transgender clients. *Journal of Voice, 27*(3), 321–334. https://doi.org/10.1016/j.jvoice.2012.07.008

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Gooren, L. J., Sungkaew, T., Giltay, E. J., & Guadamuz, T. E. (2015). Cross-sex hormone use, functional health and mental well-being among transgender men (Toms) and transgender women (Kathoeys) in Thailand. *Culture, Health & Sexuality, 17*(1), 92–103. https://doi.org/10.1080/13691058.2014.950982

Gorton, R. N., & Erickson-Schroth, L. (2017). Hormonal and surgical treatment options for transgender men (female-to-male). *Psychiatric Clinics of North America, 40*(1), 79–97. https://doi.org/10.1016/j.psc.2016.10.005

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

Hancock, A., Colton, L., & Douglas, F. (2014). Intonation and gender perception: Applications for transgender speakers. *Journal of Voice, 28*(2), 203–209. https://doi.org/10.1016/j.jvoice.2013.08.009

Hancock, A. B., Krissinger, J., & Owen, K. (2011). Voice perceptions and quality of life of transgender people. *Journal of Voice, 25*(5), 553–558. https://doi.org/10.1016/j.jvoice.2010.07.013

Hancock, A. B., & Pool, S. F. (2017). Influence of listener characteristics on perceptions of sex and gender. *Journal of Language and Social Psychology, 36*(5), 599–610. https://doi.org/10.1177/0261927X17704460

Hardy, T. L. D., Boliek, C. A., Wells, K., Dearden, C., Zalmanowitz, C., & Rieger, J. M. (2016). Pretreatment acoustic predictors of gender, femininity, and naturalness ratings in individuals with male-to-female gender identity. *American Journal of Speech-Language Pathology, 25*(2), 125–137. https://doi.org/10.1044/2015_AJSLP-14-0098

Hardy, T. L. D., Rieger, J. M., Wells, K., & Boliek, C. A. (2020). Acoustic predictors of gender attribution, masculinity–femininity, and vocal naturalness ratings amongst transgender and cisgender speakers. *Journal of Voice, 34*(2), 300.e11–300.e26. https://doi.org/10.1016/j.jvoice.2018.10.002

Henton, C. G. (1989). Fact and fiction in the description of female and male pitch. *Language & Communication, 9*(4), 299–311. https://doi.org/10.1016/0271-5309(89)90026-8

Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research, 15*(1), 155–159. https://doi.org/10.1044/jshr.1501.155

Holmberg, E. B., Oates, J., Dacakis, G., & Grant, C. (2010). Phonetograms, aerodynamic measurements, self-evaluations, and auditory perceptual ratings of male-to-female transsexual voice. *Journal of Voice, 24*(5), 511–522. https://doi.org/10.1016/j.jvoice.2009.02.002

Houle, N., & Levi, S. V. (2019). Effect of phonation on perception of femininity/masculinity in transgender and cisgender speakers. *Journal of Voice.* Advance online publication. https://doi.org/10.1016/j.jvoice.2019.10.011

Ingham, R. J., & Packman, A. (1978). Perceptual assessment of normalcy of speech following suttering therapy. *Journal of Speech and Hearing Research, 21*(1), 63–73. https://doi.org/10.1044/jshr.2101.63

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

Junger, J., Habel, U., Bröhr, S., Neulen, J., Neuschaefer-Rube, C., Birkholz, P., Kohler, C., Schneider, F., Derntl, B., & Pauly, K. (2014). More than just two sexes: The neural correlates of voice gender perception in gender dysphoria. *PLOS ONE, 9*(11) https://doi.org/10.1371/journal.pone.0111672

Junger, J., Pauly, K., Bröhr, S., Birkholz, P., Neuschaefer-Rube, C., Kohler, C., Schneider, F., Derntl, B., & Habel, U. (2013). Sex matters: Neural correlates of voice gender perception. *NeuroImage, 79,* 275–287. https://doi.org/10.1016/j.neuroimage.2013.04.105

Kawitzky, D., & McAllister, T. (2020). The effect of formant biofeedback on the feminization of voice in transgender women. *Journal of Voice, 34*(1), 53–67. https://doi.org/10.1016/j.jvoice.2018.07.017

Kent, R., & Vorperian, H. (2018). Static measurements of vowel formant frequencies and bandwidths: A review. *Journal of Communication Disorders, 74,* 74–97. https://doi.org/10.1016/j.jcomdis.2018.05.004

King, R. S., Brown, G. R., & McCrea, C. R. (2012). Voice parameters that result in identification or misidentification of biological gender in male-to-female transgender veterans. *International Journal of Transgenderism, 13*(3), 117–130. https://doi.org/10.1080/15532739.2011.664464

Kleinschmidt, D. F., & Jaeger, F. T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*(2), 148–203. https://doi.org/10.1037/a0038695

Klopfenstein, M. (2015). Relationship between acoustic measures and speech naturalness ratings in Parkinson's disease: A within-speaker approach. *Clinical Linguistics & Phonetics, 29*(12), 938–954. https://doi.org/10.3109/02699206.2015.1081293

Lane, H., Catania, A., & Stevens, S. (1961). Voice level: Autophonic scale, perceived loudness, and effects of sidetone. *The Journal of the Acoustical Society of America, 33*(2), 160–167. https://doi.org/10.1121/1.1908608

Lass, N. J., Almerino, C. A., Jordan, L. F., & Walsh, J. M. (1980). The effect of filtered speech on speaker race and sex identifications. *Journal of Phonetics, 8*(1), 101–112. https://doi.org/10.1016/S0095-4470(19)31445-7

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *The Journal of the Acoustical Society of America, 59*(3), 675–678. https://doi.org/10.1121/1.380917

Lecanuet, J.-P., Granier-Deferre, C., Jacquet, A., Capponi, I., & Ledru, L. (1993). Prenatal discrimination of a male and a female voice uttering the same sentence. *Early Development and Parenting, 2*(4), 217–228. https://doi.org/10.1002/edp.2430020405

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

Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., & Miller, M. (1985). Measures of the sentence intonation of read and spontaneous speech in American English. *The Journal of the Acoustical Society of America, 77*(2), 649–657. https://doi.org/10.1121/1.391883

Lotfian, R., & Busso, C. (2017). Building naturalistic emotionally balanced speech corpus by retrieving emotional speech from existing podcast recordings. *Transactions on Affective Computing, 10*(4), 471–483. https://doi.org/10.1109/TAFFC.2017.2736999

Mackey, L. S., Finn, P., & Ingham, R. J. (1997). Effect of speech dialect on speech naturalness ratings: A systematic replication of Martin, Haroldson, and Triden (1984). *Journal of Speech, Language, and Hearing Research, 40*(2), 349–360. https://doi.org/10.1044/jslhr.4002.349

Metz, D. E., Schiavetti, N., & Sacco, P. R. (1990). Acoustic and psychophysical dimensions of the perceived speech naturalness of nonstutterers and posttreatment stutterers. *Journal of Speech and Hearing Disorders, 55*(3), 516–525. https://doi.org/10.1044/jshd.5503.516

Miller, C. L., Younger, B. A., & Morse, P. A. (1982). The categorization of male and female voices in infancy. *Infant Behavior and Development, 5*(2–4), 143–159. https://doi.org/10.1016/S0163-6383(82)80024-6

Mount, K. H., & Salmon, S. J. (1988). Changing the vocal characteristics of a postoperative transsexual patient: A longitudinal study. *Journal of Communication Disorders, 21*(3), 229–238. https://doi.org/10.1016/0021-9924(88)90031-7

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation.

Language and Speech, 50(1), 125–142. https://doi.org/10.1177/00238309070500010601

Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language, 22*(2), 171–184. https://doi.org/10.1016/j.csl.2007.07.003

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

Oates, J. M., & Dacakis, G. (2015). Transgender voice and communication: Research evidence underpinning voice intervention for male-to-female transsexual women. *SIG 3 Perspectives on Voice and Voice Disorders, 25*(2), 48–58. https://doi.org/10.1044/vvd25.2.48

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51,* 195–203. https://doi.org/10.3758/s13428-018-01193-y

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24*(2), 175–184. https://doi.org/10.1121/1.1906875

Pettit, J. M. (2004). Transsexualism and sex reassignment: Speech differences. In R. D. Kent (Ed.), *The MIT encyclopedia of communication disorders* (pp. 223–225). MIT Press.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics, 2,* 33–52. https://doi.org/10.1146/annurev-linguistics-030514-125050

Queen, R. M. (1998). 'Stay queer!' 'Never fear!': Building queer social networks. *World Englishes, 17*(2), 203–214. https://doi.org/10.1111/1467-971X.00094

Remez, R. E., Rubin, P. E., & Ball, S. (1985). Sentence intonation in spontaneous utterances and fluently spoken text. *The Journal of the Acoustical Society of America, 77*(S1), S38. https://doi.org/10.1121/1.2022306

Roenfeldt, K. (2018). *Better than average: Calculating geometric means using SAS* (pp. 1–9). Henry M. Jackson Foundation for the Advancement of Military Medicine. https://www.lexjansen.com/wuss/2018/56_Final_Paper_PDF.pdf

Scheidt, D., Kob, M., Willmes, K., & Neuschaefer-Rube, C. (2004, January). *Do we need voice therapy for female-to-male transgenders?* Paper presented at the 26th World Congress of the IALP, Brisbane, Australia.

Schilt, K., & Westbrook, L. (2009). Doing gender, doing heteronormativity: "Gender normals," transgender people, and the social maintenance of heterosexuality. *Gender & Society, 23*(4), 440–464. https://doi.org/10.1177/0891243209340034

Schwartz, M. (1968). Identification of speaker sex from isolated, voiceless fricatives. *The Journal of the Acoustical Society of America, 43*(5), 1178–1179. https://doi.org/10.1121/1.1910954

Schwartz, M., & Rine, H. (1968). Identification of speaker sex from isolated, whispered vowels. *The Journal of the Acoustical Society of America, 44*(6), 1736–1737. https://doi.org/10.1121/1.1911324

Schwarz, K., Fontanari, A. M. V., Costa, A. B., Soll, B. M. B., da Silva, D. C., de Sá Villas-Bôas, A. P., Cielo, C. A., Bastilha, G. R., Ribeiro, V. V., Dorfman, M. E. K. Y., & Lobato, M. I. R. (2018). Perceptual–auditory and acoustical analysis of the voices of transgender women. *Journal of Voice, 32*(5), 602–608. https://doi.org/10.1016/j.jvoice.2017.07.003

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

Smith, E., Junger, J., Pauly, K., Kellermann, T., Neulen, J., Neuschaefer-Rube, C., Derntl, B., & Habel, U. (2018). Gender incongruence and the brain—Behavioral and neural correlates of voice gender perception in transgender people. *Hormones and Behavior, 105,* 11–21. https://doi.org/10.1016/j.yhbeh.2018.07.001

Snow, M. P., & Williges, R. C. (1998). Empirical models based on free-modulus magnitude estimation of perceived presence in virtual environments. *Human Factors, 40*(3), 386–402. https://doi.org/10.1518/001872098779591395

Southwood, M. (1996). Direct magnitude estimation and interval scaling of naturalness and bizarreness of the dysarthria associated with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology, 4*(1), 13–25. https://www.dds.indiana.edu/illiad/illiad.dll?Action=10&Form=75&Value=1870789

Stevens, S. (1975). *Psychophysics*. Wiley. https://doi.org/10.1111/j.1469-8986.1975.tb00006.x

Stoicheff, M. (1981). Speaking fundamental frequency characteristics of nonsmoking female adults. *Journal of Speech and Hearing Research, 24*(3), 437–441. https://doi.org/10.1044/jshr.2403.437

Strand, E. (2000). *Gender stereotype effects in speech process-ing* [Unpublished doctoral dissertation]. The Ohio State University.

Talaat, M., Angelo, A., Talaat, A. M., Elwany, S., Kelada, I., & Thabet, H. (1987). Histologic and histochemical study of effects of anabolic steroids on the female larynx. *Annals of Otology, Rhinology & Laryngology, 96*(4), 468–471. https://doi.org/10.1177/000348948709600423

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

van Borsel, J., de Cuypere, G., & Van den Berghe, H. (2001). Physical appearance and voice in male-to-female transsexuals. *Journal of Voice, 15*(4), 570–575. https://doi.org/10.1016/S0892-1997(01)00059-5

van Borsel, J., Vandaele, J., & Corthals, P. (2013). Pitch and pitch variation in lesbian women. *Journal of Voice, 27*(5), 656.e13–656.e16. https://doi.org/10.1016/j.jvoice.2013.04.008

van Son, R. J. J. H. (2005). A study of pitch, formant, and spectral estimation errors introduced by three lossy speech compression algorithms. *Acta Acustica United With Acustica, 91*(4), 771–778. https://hdl.handle.net/11245/1.247267

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58,* 475–482. https://doi.org/10.1007/BF03395630

Wolfe, V. I., & Ratusnik, D. L. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorders, 55*(1), 43–50. https://doi.org/10.1044/jshd.5501.43

Wong, P., & Babel, M. (2017). Perceptual identification of talker ethnicity in Vancouver English. *Journal of Sociolinguistics, 21*(5), 603–628. https://doi.org/10.1111/josl.12264

Yoho, S. E., Borrie, S. A., Barrett, T. S., & Whittaker, D. B. (2019). Are there sex effects for speech intelligibility in American English? Examining the influence of talker, listener, and methodology. *Attention, Perception, and Psychophysics, 81*(2), 558–570. https://doi.org/10.3758/s13414-018-1635-3