PERCEIVED GENDER IN CLEAR AND CONVERSATIONAL SPEECH

by

Jaime A. Booz

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Communication Sciences and Disorders

The University of Utah

December 2016

b. 11/22/1984

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The thesis of **Jaime A. Booz 11/22/1984**

has been approved by the following supervisory committee members:

Sarah Hargus Ferguson	, Chair	**05/06/2016** Date Approved
Bruce Smith	, Member	**05/06/2016** Date Approved
Julie M. Barkmeier-Kraemer	, Member	**05/06/2016** Date Approved

and by **Michael Blomgren** , Chair/Dean of

the Department/College/School of **Communication Sciences and Disorders**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Although many studies have examined acoustic and sociolinguistic differences between male and female speech, the relationship between talker speaking style and perceived gender has not yet been explored. The present study attempts to determine whether clear speech, a style adopted by talkers who perceive some barrier to effective communication, shifts perceptions of femininity for male and female talkers.

Much of our understanding of gender perception in voice and speech is based on sustained vowels or single words, eliminating temporal, prosodic, and articulatory cues available in more naturalistic, connected speech. Thus, clear and conversational sentence stimuli, selected from the 41 talkers of the Ferguson Clear Speech Database (Ferguson, 2004) were presented to 17 normal-hearing listeners, aged 18 to 30. They rated the talkers' gender using a visual analog scale with "masculine" and "feminine" endpoints. This response method was chosen to account for within-category shifts of gender perception by allowing nonbinary responses.

Mixed-effects regression analysis of listener responses revealed a small but significant effect of speaking style, and this effect was larger for male talkers than female talkers. Because of the high degree of talker variability observed for talker gender, acoustic analyses of these sentences were undertaken to determine the relationship between acoustic changes in clear and conversational speech and perceived femininity. Results of these analyses showed that mean fundamental frequency (f_0) and f_0 standard deviation were significantly correlated to perceived gender for both male and female talkers, and vowel space was significantly correlated only for male talkers. Speaking rate and breathiness measures (CPPS) were not significantly related for either group.

Outcomes of this study indicate that adopting a clear speaking style is correlated with increases in perceived femininity. Although the increase was small, some changes associated with making adjustments to improve speech clarity have a larger impact on perceived femininity than others. Using a clear speech strategy alone may not be sufficient for a male speaker to be perceived as female, but could be used as one of many tools to help speakers achieve more "feminine" speech, in conjunction with more specific strategies targeting the acoustic parameters outlined in this study.

CONTENTS

LIST OF FIGURES

Figures

LIST OF TABLES

Tables

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Introduction

The speech of a particular talker carries not only the linguistic message, arguably the purpose of verbal interaction, but also a number of indexical characteristics. We as listeners hear not only what our communication partner attempts to convey through words, but also information about their gender, age, dialect, race, sexual orientation, emotional state, intelligence, and any number of other characteristics. These perceptions influence judgments we make about the speaker and, consequently, the way we respond. In short, the unspoken message plays as much a part in human communication as the spoken message.

While listeners adjust their assumptions about talkers based on both spoken and unspoken information, talkers also adjust their speech based on what they know or assume about the listener. This is obvious in the case of child-directed speech, in which a speaker simplifies their language and raises their pitch. The present study is concerned with the clear speaking style. This style is adopted by talkers when they either know or judge that their listener does not understand or cannot hear them, such as if the listener has a hearing loss or is not a fluent speaker of the talker's native language. Although clear speech strategies can positively impact intelligibility, making these adjustments can also influence perception of speaker indexical characteristics. This study aims to examine whether clear speech strategies influence perception of talker gender.

If adopting a clear speaking style does indeed impact perception of gender, then a talker could purposely use this method to influence how their gender is perceived in their communication partners. This is of particular interest to transgender individuals, and especially to transgender women, who may desire treatment to achieve a voice and communication style that more closely aligns with their gender expression. Additionally, there are pathological voice conditions that alter pitch and thus may influence listener perceptions of gender, especially when no visual information is present, such as during telephone conversations. Because a clear speaking style can be easily elicited by simply instructing someone to talk as though communicating with someone who does not understand them, results of this study could inform strategies to efficiently alter gender presentation through voice.

Gender Differences in Voice and Speech

Anatomical Differences

The sexual dimorphism of male and female laryngeal anatomy occurs during and after puberty, concurrent with other muscle and bone growth in the body. Growth of the thyroid and cricoid cartilages occurs in both males and females, but it is 2-3 times greater in males, both in dimensions and in weight (Kahane, 1982). Additionally, these cartilages maintain their proportions in both sexes during growth, with the exception of the anterior-posterior dimension of the thyroid cartilage in males, which increases some three times more than that of females (Kahane, 1982; Titze, 1989). This growth increases the distance between the vocal process and the anterior commissure, which disproportionately lengthens male vocal folds.

In concert with the growth of the laryngeal cartilages, the vocal fold tissue also grows. Male vocal folds increase in length by 63%, whereas female vocal folds increase only

34% (Kahane, 1982; Kazarian, 1978; Titze, 1989). Typical male vocal folds measure 16 mm in length while female vocal folds average 10 mm in length (Titze, 1989). Although the longer male vocal folds account for a lower fundamental frequency (f_o), male vocal folds are also typically thicker than those of females. The enlarged thyroarytenoid muscle achieves a more complete closure during phonation, and thus a longer closed period during each cycle of vibration. This allows a greater proportion of the cover of the vocal fold to vibrate, and also yields a greater amplitude of vibration. Consequently, thicker vocal folds produce a different vocal quality in the male voice (Titze, 1994). Pubertal growth of laryngeal structures is largely attributed to testosterone, although female vocal folds are also subject to hormonal influence. Further changes in vocal fold tissue occur in aging. Male membranous vocal folds shorten, female vocal fold mucosa and cover thicken, and both sexes experience edema in the superficial lamina propria (Hirano, Kurita, & Sakaguchi, 1989).

Female and male oral and pharyngeal cavities are also dimorphic. The male vocal tract is typically 15% longer, though males and females have roughly equivalent oral tract length. The difference in pharyngeal tract length thus accounts for the discrepancy (Goldstein, 1980): females typically have longer oral cavities than pharyngeal cavities, while the opposite is true for males. Differences in vocal tract dimensions are thought to account for differences in formant frequencies.

Acoustic Differences

The physical differences between male and female vocal mechanisms result in some notable acoustic differences. Fundamental frequency (f_o), associated with the rate of oscillation of the vocal folds, is slightly less than an octave lower in adult males than in females as a result of differences in length and weight of the membranous portion of the vocal folds. The average male f_o is around 130 Hz, whereas the average female's is around

220 Hz (Hillenbrand, Getty, Clark, & Wheeler, 1995). Resonance, associated with the vibration of air in the supralaryngeal vocal tract, is affected by the size and shape of the vocal tract as well as by modification of this size and shape by placement of the lips and tongue. As males have a longer vocal tract, their formant frequencies are lower (Bachorowski & Owren, 1999), though only by a scale of 1.18, 1.17, and 1.14 for F1, F2, and F3, respectively (Hillenbrand et al., 1995). Neel (2008) examined vowel space for normal male and female talkers and found a greater vowel perimeter for female talkers (25.07 Barks versus 18.57 Barks).

Although these anatomical differences create seemingly dissimilar acoustic features, there is a great deal of overlap between male and female talkers depending on individual speaker, context, and speaking task (Maurer & Landis, 1996; Peterson & Barney, 1952). Measurements of f_o of an isolated vowel, for example, do not give the same acoustic picture as measurements of average f_o or f_o range across a sentence production task, where f_o changes coincide with linguistic and paralinguistic features. It should also be noted that f_o is relatively independent of phonemic context, while formant frequencies change considerably with relation to the speech sound being produced.

In addition to fundamental and formant frequencies, other glottal characteristics of female physiology cause differences in voice quality. It is common for female speakers to have a posterior glottal opening which persists throughout approximation of the vocal folds, something that is not seen frequently in male speakers (Sodersten & Lindestad, 1990). Holmberg, Hillman, and Perkell (1988) also found that females typically had a larger open quotient at normal loudness, which is a ratio of the time the vocal folds are open to the total time of one glottal vibratory cycle. They also had more gradual rises and falls in glottal flow. These varying glottal characteristics correlate with greater levels of aperiodic noise around

F3, reduced amplitude of higher-frequency harmonics, and thus higher relative amplitude of F1. These acoustic measures together result in steeper spectral tilt and a "breathy" voice quality. Cepstral Peak Prominence Smoothed (CPPS) is the acoustic measure that is most highly correlated with perception of breathy voice quality (Lowell, Colton, Kelley, & Mizia, 2013). This measure is calculated logarithmically and is, essentially, a spectrum of the speech signal spectrum. This measure is sensitive to breathiness in connected speech samples, but is also affected by roughness and glottal fry. Time-based measures of voice quality deviance, such as jitter and shimmer, can only be used with sustained phonation and are susceptible to high amounts of aperiodic noise that may mask the signal. CPPS is largely used for evaluation of dysphonic voices; however, norms were developed by Garrett (2013). Awan, Giovinco, and Owens (2011) suggesting that CPPS is also sensitive to changes in intensity and vowel, with increases in CPPS (decreased breathy vocal quality) related to increases in intensity. Their study also reported higher mean CPPS values for males than females, consistent with the assertion that female voices are more breathy.

Perception of Gender in Speech

It should be noted, at this juncture, that perception of speaker sex and speaker gender may not be congruent if the talker is transgender. This study is interested not in perceptions of talker sex, or results of the talker's vocal physiology alone, but in talker gender, which includes behavioral as well as physical attributes of the speaker. Listeners are generally very accurate in their judgments of a speaker's gender based on their voice alone, but it is not entirely known which acoustic properties precisely play a role in shifting perceptions of gender. Lower f_0 and formant frequencies correspond to longer vocal folds and longer vocal tracts respectively, and thus a larger speaker. The presence of the posterior glottic gap in females increases aperiodic noise, so perception of breathiness also signals some physical property in the speaker. Listeners are even relatively accurate at identifying talker sex in whispered vowels, where f_0 information is not available (Schwartz & Rine, 1968). While listeners are fairly accurate at identifying speaker sex in vowel segments in both natural and synthesized speech, modifying f_0 or formant frequencies separately reduces their accuracy (Hillenbrand & Clark, 2009). Because segmented vowels or sustained phonation cannot capture temporal or prosodic variation between or within talkers, and because so many other social factors are at play in identification of talker sex, shifts in perception cannot be accounted for by changes in f_0, formant frequencies, and voice quality alone. Further, studies that offer only two options for a listener to choose from will accurately measure listeners' ability to identify talker sex, but may fail to capture within-category differences in perceived gender.

A male talker who raises his pitch is unlikely to achieve a f_0 perceived as feminine, but the range between 150 and 180 is often perceived as gender ambiguous. Because f_0 seems to be the dominant gender cue, it may be more difficult to influence listener perceptions of femininity in male voices than the reverse. One possible explanation for this is that male voices are a divergence from what would have been a shared vocal change trajectory with female voices. This makes male voice the "marked" form, and will ensure that speakers within a male f_0 range will almost always be perceived as adult males (Hillenbrand, 2009; Owren, Berkowitz, & Bachorowski, 2007).

Sociolinguistic Differences

A number of socially influenced language and pragmatic differences exist between male and female speakers. However, when studying perception of gender in the voice, most studies neutralize this effect by having talkers read passages aloud or using stimuli with only sustained vowels. As sustained phonation does not mimic actual speech production, reading of neutral sentences or passages may be the best way to achieve gender-based prosody patterns without introducing lexical and social gender cues to the stimuli. Perception of femininity in a speaker may be shifted to some extent by these socially-influenced acoustic cues of gender.

First, although male and female voices have an average range of possible f_os, the range they actually use may be influenced by social factors. Perry, Ohde, and Ashmead (2001) found that listeners are able to identify talker sex from recordings of children's voices with high accuracy, despite the fact that prepubertal males should have similar f_o to pre-pubertal females. This implies that males and females may utilize their f_o ranges differently based on social cues, even in childhood.

Further, f_o variability may be associated with perceived femininity. Female talkers typically use more exaggerated prosody (Freidenberg, 2002; Oates & Dacakis, 1997), which translates to more voices that have greater f_o range and a higher upper limit of f_o being judged as more feminine, even when the speaker has a physiologically male vocal tract (Gelfer & Schofield, 2000). Female speakers also tend to use rising final intonation more often (Freidenberg, 2002).

Clear Speech

Talkers make any number of speech adjustments to accommodate to environments, listeners, or their own internal state. Clear speech encompasses the adjustments that talkers make to their speaking style when their communication partner has hearing loss or otherwise has difficulty understanding them. This can be elicited in a reading task by asking talkers to speak as though talking to someone with a hearing loss.

The acoustic characteristics of clear speech are quantitatively measured relative to those of casual speech under typical speaking conditions, commonly referred to in the literature as conversational speech. Talkers using clear speech speak with a slower speaking rate (Bradlow, 2003; Ferguson et al, 2010; Picheny, Durlach, & Braida et al, 1986), raise their average f_o (Bradlow et al, 2003; Hazan & Baker, 2011), use more f_o variability (Bradlow et al, 2003; Hazan & Baker, 2011; Picheny et al, 1986), and increase their loudness (Ferguson et al, 2010). In addition, speakers make specific adjustments to vowels in clear speech, namely, producing an expanded vowel space, longer vowel duration, and greater dynamic formant movement (Ferguson & Kewley-Port, 2002, 2007; Ferguson & Quené, 2014; Picheny et al, 1986). These changes are associated with opening the mouth wider, using more extreme articulatory positions of vowels (e.g., more back low/back vowels and more fronted front vowels) and increasing vocal effort. Specific adjustments to consonants include stronger final consonants (Bradlow et al., 2003; Ferguson et al., 2010; Picheny et al., 1986) and longer fricative duration (Maniwa, Jongman, & Wade, 2009). While talkers make these adjustments almost unconsciously in response to adverse listening conditions, the magnitude of these changes varies widely among talkers (Ferguson & Quené, 2014). Ferguson and Kewley-Port (2007) found that speakers that had a bigger difference between their clear and conversational speech (more clear speech benefit) showed a bigger difference in vowel space and vowel duration between the two speaking styles than those that did not. However, hearing loss may impact which acoustic cues are important in vowel intelligibility (Ferguson & Quené, 2014).

Similarities Between Feminine Speech and Clear Speech

Because some of the acoustic changes associated with increased intelligibility in clear speech overlap with characteristics of speech perceived as feminine, presumably clear speech strategies will also increase perception of femininity of the speaker. Most notably, the raised f_o, expanded vowel space, longer vowel duration, and greater f_o variability observed in clear speech overlap with typical acoustic differences between male and female speech. However, due to the increased intensity and vocal effort associated with clear speech, talkers may have reduced breathiness and thus increased CPPS. Increased loudness and reduced breathy vocal quality may correlate more with masculine perception of voice in clear speech as a result. Table 1.1 summarizes these and other similarities and differences.

Hypotheses

This study builds, then, on three key pieces of information related to clear speech and gender perception: first, that clear speech shares some important acoustic features associated with perception of speech as feminine; second, that perception of gender is not categorical and that allowing more than two binary options will provide more information about which acoustic features correspond with shifts in listener judgments of femininity in clear speech stimuli; and third, that there is a great degree of variability in how talkers produce clear speech. With the above points in mind, the hypotheses of this study are as follows:

- Clear speech will be perceived as more feminine across talkers than conversational speech.
- Perception of femininity will shift more in clear speech for talkers who produce clear speech with a higher f_0 and a higher f_0 variation than for talkers who do not produce these changes.

With regards to breathiness, two possible hypotheses exist:

- Breathy vocal quality may be reduced in female talkers and thus may result in reduced perception of femininity in clear speech, or
- Breathy vocal quality will not play a part in judgments of femininity in clear speech.

Table 1.1

Acoustic Correlates of Clear Speech Compared to Speech Perceived as Feminine

Clear speech	**Feminine Speech**
Higher f_o (men 6 Hz, women 2 Hz; Ferguson, Morgan, Rogers, & Hunter, 2014)	Higher f_o
Higher F1 and F2	Higher F1 and F2
More expanded vowel space	More expanded vowel space (derived from Hillenbrand et al. 1995)
More pitch variability, more for male speakers (Bradlow 2003)	More pitch variability (Freidenberg, 2002; Oates & Dacakis, 1997)
Probably higher CPPS due to increased vocal effort	Lower CPPS
More fronted articulation for front vowels	More fronted articulation overall
Longer duration	Longer duration (Hillenbrand et al., 1995)

CHAPTER 2

METHODS

Review of Current Literature

No study exists that examines the effect of speaking style changes on perceived femininity. A large number of studies have examined the perception of gender in normal speech of men, women and children, transgender speech, and synthesized speech, with varying results.

Hillenbrand and Clark (2009) examined the intercorrelation of f_0 and formant frequencies on perceived gender using synthesized speech. Fifty talkers (25 male, 25 female) were recorded reading both sentences and carrier phrases. Stimuli consisted of the sentences and vowels excised from the carrier phrases, presented with either modulated f_0, modulated formant frequencies, or both. Listeners made binary judgments, as well as providing a confidence rating, meant to capture participants' uncertainty about their ratings. Ultimately, shifting both formant frequencies and f_0 had more effect than shifting one parameter alone. While synthesized speech allows manipulation of individual speech parameters, it sacrifices speech naturalness as a result. In addition, restricting listeners to binary responses fails to capture within-category shifts.

Backarowski and Owren (1999) similarly found that identification of talker sex was most accurate when f_0 and formant frequencies were congruent for the talker's sex. This study used 2500 isolated vowel segments and required listeners to decide whether the talker was male or female. Similar to Hillenbrand and Clark's (2009) study, this experiment focused on accuracy of talker sex identification, rather than looking at acoustic features that may cause listeners to perceive smaller, within-category perceptual changes. Regardless of the limitations of using binary choice and vowel-only stimuli, both of the above studies showed that both f_o and formant frequencies are an important parameter in listener's perception of talker sex.

Several other recent perceived gender experiments used a Likert scale response method, giving listeners the ability to rate speakers' femininity or masculinity rather than sex. Honorof and Whalen (2010) presented sustained vowels at extremes of talkers' f_o ranges and asked listeners to rate what they heard from "male for sure" to "female for sure." Using "male" and "female" as endpoints still encourages listeners to rate the talker's sex rather than to listen for smaller, more nuanced shifts in perceived gender. Results showed that listeners typically perceived higher f_os as female, regardless of talker sex, consistent with other studies, which showed high correlation between higher f_o and perception of femininity. Further, we can infer that male talkers can achieve a f_o that results in being perceived as female; however, sustained phonation does not capture prosodic or temporal cues. Additionally, f_o extremes are not an ecologically valid representation of the f_o a talker would typically use, nor is it clinically applicable, as using the highest achievable f_o is not a sustainable use of the vocal mechanism.

Similar to the above study, Wolfe, Ratusnik, Smith, and Northrop (1990) found that f_o was the most salient cue of talker femininity for transgender talkers. This study used short narrative samples from transgender and cisgender (those whose gender identity and gender assigned at birth are congruent, i.e., the opposite of transgender) talkers as stimuli, which listeners rated on a seven-point Likert scale from "extremely feminine" to "extremely masculine." Narrative samples allowed listeners to use prosodic and temporal cues to make their judgments. As a result, prosody did influence perception of femininity; specifically, talkers who were rated as more feminine used more upward intonation patterns. There is a possibility that sociolinguistic variables are introduced in narrative-level information, especially given that this study used personal narrative rather than a structured narrative task, such as picture description.

Ko, Judd, and Blair (2006) reported similar findings with regard to f_o variability. Their study aimed to determine which vocal cues influenced within-category gender judgments. This three-part study used Rainbow Passage recordings from 94 talkers presented in three different contexts. Results showed that f_o, formant frequencies, and f_o variability played a large role in listener's determination of gender category, but that f_o variability, measured as f_o standard deviation, was more significantly correlated with femininity shifts in male than female talkers. Interestingly, f_o alone was correlated with shifts in perceived femininity for female talkers.

Few studies have looked at vocal quality and perception of femininity. Van Borsel, Janssens, and de Bodt (2007) trained female speech-language pathology students to produce sustained vowels with varying degrees of breathy vocal quality. These stimuli were used for each of two experiments. The first asked listeners to rate the stimuli on Likert scale from "little feminine" to "very feminine." This selection is appropriate for a talker pool of only female talkers, but still allows for within-category shifts of perceived femininity. The second experiment presented two stimuli pairwise and asked listeners to rate which sounded more feminine. Results of both studies showed that breathiness was correlated with increased perceived femininity, even within the same talker. Although the experiments used only sustained phonation, this and a similar study using word- or sentence-level materials might yield more information about the impact of breathiness on perceived femininity.

It is important to point out that the majority of perceived gender studies examine perceived femininity rather than masculinity. While this is perhaps more salient in determining clinical utility of these studies to the transgender speaker, as transgender men accomplish f_0 changes through hormones rather than behavioral changes, and are thus less likely to seek voice and communication training, some behavioral changes may still impact perceived masculinity. Avery and Liss (1996) examined the acoustic correlates of perceived masculinity in male speakers. Female listeners were presented with connected speech samples of a "baseline" talker, and asked to rate the femininity of a second sample in comparison to the baseline. Results revealed that not only was f_0 variation important, but the pattern of f_0 contour was also significant in shifting perception of masculinity. In addition, examination of vowel formants showed less reduction in the less-masculine-sounding speakers, suggesting that they produced more clear speech than the more-masculine-sounding speakers.

Participants

Seventeen subjects (eight female, nine male) were recruited either from the University of Utah Psychology Department pool or via in-class announcement in University of Utah Communication Sciences and Disorders (CSD) undergraduate classes. Participants ranged in age from 18 to 30, (M = 22.6). All but three participants were from Utah, with the three remaining native to northern California, Colorado, and Nebraska. Demographic data for all participants are shown in Table 2.1. All reported that they spoke English as their first language and denied history of hearing loss or speech or language problems or therapy. Psychology department pool participants received research credit for their participation. Subjects recruited from CSD classes were paid for their participation.

Table 2.1

Demographic Information for All Participants

Code	Sex	Age	Home State	Race/Ethnicity
F01	F	18	Utah	White/Not Hispanic
F02	F	21	Colorado	Hispanic
F03	F	19	Utah	White/Not Hispanic
F04	F	19	Utah	White/Not Hispanic
F05	F	27	Utah	White/Not Hispanic
F06	F	23	Utah	White/Not Hispanic
F07	F	22	Utah	White/Not Hispanic
F08	F	21	Utah	White/Not Hispanic
M01	M	25	Utah	White/Not Hispanic
M02	M	24	Utah	White/Not Hispanic
M03	M	24	Utah	White/Not Hispanic
M04	M	22	Utah	White/Not Hispanic
M05	M	21	Utah	White/Not Hispanic
M06	M	26	Idaho	White/Not Hispanic
M07	M	20	Utah	White/Not Hispanic
M08	M	25	Nebraska	White/Not Hispanic
M09	M	30	California	White/Not Hispanic
Mean		22.6		
Standard Deviation		3.08		

Materials

This study utilized sentence materials from the Ferguson Clear Speech Database (2004). This database includes recordings of 41 talkers, five male and five female from each of four age categories: 18-24, 25-31, 32-38, and 39-45, plus an additional female talker in the 18-24 age category. Talker demographic data are reported in Table 2.2 (Ferguson, 2002). Each talker was recorded reading aloud a list of 188 sentences in two speaking styles. The list was comprised of the following: neutral carrier sentence frames containing a /bVd/ test word with one of 10 vowels; neutral carrier sentences containing a consonant-vowel-consonant word chosen from the Northwestern University Auditory Test No. 6 (NU-6; Tillman & Carhart, 1966); and selected sentences from the Central Institute for the Deaf (CID) Everyday Sentences test (Davis & Silverman, 1978). This study used only the /bVd/ sentences, so only descriptions of these sentences are included here. For each of the ten vowels (/i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u/), each talker recorded seven tokens (each in a different sentence frame) in each speaking style, totaling 140 /bVd/ sentences (70 per speaking style). Recordings were made in a sound-attenuating booth using a headset microphone. In the conversational condition, talkers were instructed to read the sentences as they would in their everyday speaking style. In the clear speech condition, they were instructed to "speak clearly, so that a hearing impaired person would be able to understand you" (Ferguson, 2004, p. 2366). Talkers were given a set of sentences in each speaking style to practice before beginning recording.

While these sentence stimuli are read rather than spontaneous, they are closer to a talker's natural production than isolated vowels or words, which many gender perception experiments have used. Using connected speech allows f_0 variability to be measured as a potential influencing acoustic property. Sentence frames were selected based on a number of

Table 2.2

Demographic Information for All Talkers

From the Ferguson Clear Speech Database (Ferguson 2002)

Code	Sex	Age	Code	Sex	Age
F01	F	44	M01	M	21
F02	F	25	M02	M	23
F03	F	20	M03	M	20
F04	F	20	M04	M	21
F05	F	21	M05	M	22
F06	F	20	M06	M	44
F07	F	21	M07	M	37
F08	F	22	M08	M	45
F09	F	29	M09	M	35
F10	F	27	M10	M	37
F11	F	37	M11	M	38
F12	F	26	M12	M	41
F13	F	29	M13	M	31
F14	F	32	M14	M	27
F15	F	40	M15	M	25
F16	F	41	M16	M	26
F17	F	37	M17	M	33
F18	F	42	M18	M	28
F19	F	33	M19	M	45
F20	F	35	M20	M	41
F21	F	43			

factors. First, because the present study looks at perceived gender, any semantic cues that could influence listener perceptions of the talker's gender were neutralized by choosing sentence frames that did not contain gendered names, pronouns, or any reference to potentially gendered constructs. Second, because higher formant frequencies are an acoustic correlate of perceived femininity, there was a possibility that vowel may be a factor in perception of talker femininity in this study. Thus, selected sentences contained a balanced selection of front/back and low/high vowels. Last, given that rate measures were calculated in syllables per second, choosing sentences of equal syllable length facilitated convenient comparisons of speaking rate between different sentence frames. Eight sentences per speaking style were selected from each talker's set of /bVd/ neutral sentences, 16 sentences per talker. A list of these sentences is provided in Appendix A.

All 41 talkers were used for this study to maximize talker variability. Each listener heard and rated a total of 656 sentence stimuli in the test condition. Stimuli were arranged into eight blocks, each block containing two sentences from each talker: one clear and one conversational token of a single sentence. Thus, each block contained 82 identical sentences. There was a possibility that contrast effects would influence perceived femininity if stimuli blocks were mixed-gender (Hubbard & Assmann, 2013). To minimize these effects as much as possible, stimuli were randomized within blocks. Each listener received the blocks in a random order.

A familiarization task was also prepared using 41 sentences, one from each talker. These were selected from the CID Everyday Sentences recorded by the same talkers used in the present study (and used in Ferguson et al., 2010). Both conversational and clear productions were presented. These sentences were already clipped, silence added, and scaled to the correct average root mean square (RMS) intensity, as outlined for test stimuli below, so no adjustments were made to these stimuli prior to presentation.

Once stimuli were selected, they were prepared. Uniform periods of silence were added to the beginning and end of each stimulus using a MATLAB (The MathWorks, Inc., 2014a) script that opened each file, determined the onset and offset points using an arbitrary amplitude criterion of 0.01, found the nearest zero crossing, and clipped any extraneous sound preceding or following the sentence. The script then added 50 ms of silence to the beginning and end of each sentence.

The experimenter validated this procedure by manually clipping 20 sentence files using Cool Edit 2000 (Syntrillium Software Corporation, 2000). Each sentence was analyzed, the onset and offset were determined using both the waveform and the spectrogram. Any extraneous noise preceding and following the sentence was clipped, and 50 ms of silence was added manually to the beginning and end of the sentence. A comparison of the manual and automated procedures found that the two methods produced very similar results.

After all 656 sentence stimuli were batch-processed with the MATLAB script, each sentence was reviewed to ensure that no important acoustic information had been removed and that no extraneous noise had been left. Only sentence files that included an audible breath or nonspeech vocal tract noise were incorrectly clipped by the script. These 55 stimuli were then adjusted manually.

To remove intensity cues as a possible influencing factor in perceived femininity, all sentence stimuli were scaled to the same average RMS intensity using a custom MATLAB script.

Procedure

This study was conducted in the Speech Perception Laboratory at the University of Utah. Each participant completed two approximately 90 min sessions 4-10 days apart. When a participant arrived for their first session, consent was obtained, and then they filled out an information sheet containing demographic information, including age, gender, ethnic and racial information, and questions addressing dialect. A copy of this form can be found in Appendix B.

The familiarization task and first four blocks were presented in the first session, and the last four blocks were presented in the second session. Listeners were tested individually in a quiet room seated in front of computer monitor, keyboard, and mouse. Stimuli were presented using a custom MATLAB script. On each trial, a test sentence was played out from a Tucker Davis Technologies (TDT) RP-2 real time processor. Next, the sentences were attenuated using a TDT programmable attenuator (PA-5) to a comfortable listening level. The speech was then routed via a headphone buffer (TDT HB7) for diotic presentation via Shure studio headphones (SRH840).

Once listeners were comfortably seated, they were given the following instructions orally:

> When we hear speech, we hear more than just the words the person is saying. We also hear other things like the talker's gender, their age, or their emotional state. For this experiment you will hear a number of sentences. We would like you to rate how masculine or feminine the speaker sounds to you. We are not asking you to identify if the speaker is female or male, but rather how masculine or feminine each sentence sounds. Rate the sentences based on how the speech sounds rather than on the content of the sentences. Place the marker by clicking and dragging the slider, and then press "enter" to continue to the next sentence. You may listen to each sentence a second time by clicking the "listen again" button.

Listeners were then instructed to begin the familiarization task. After the first five trials, each listener was given the opportunity to adjust the presentation level to a more comfortable setting and to ask any questions that arose. The researcher then exited the room and monitored the listener's progress via a second computer screen for the remainder of the testing session. A break was offered after the completion of each block to minimize listener fatigue. Instructions were given again at the beginning of the second session, and listeners were offered a chance to repeat the familiarization task if desired.

On each familiarization and experiment trial, the custom MATLAB script displayed a slider bar representing a visual analog scale with "masculine" and "feminine" as its endpoints. Listeners clicked and dragged the slider bar to indicate how masculine or feminine the speaker sounded to them (see Figure 2.1). Endpoints of the response scale were alternated between feminine → masculine and masculine → feminine so that half the participants received one scale orientation and the other half received the opposite to minimize response bias. Each response was coded by the MATLAB script as a whole number between 0 and 100, which corresponded to the end points of the scale. Because higher values corresponded to the right end point of the scale, half the listener's ratings required adjustment so that higher values corresponded to femininity ratings for all listeners. For those participants who were given the feminine → masculine response orientation, each data point was subtracted from 100 so that all responses fit the masculine → feminine response orientation. In other words, each data point was the listener's rating of how feminine each talker sounded.

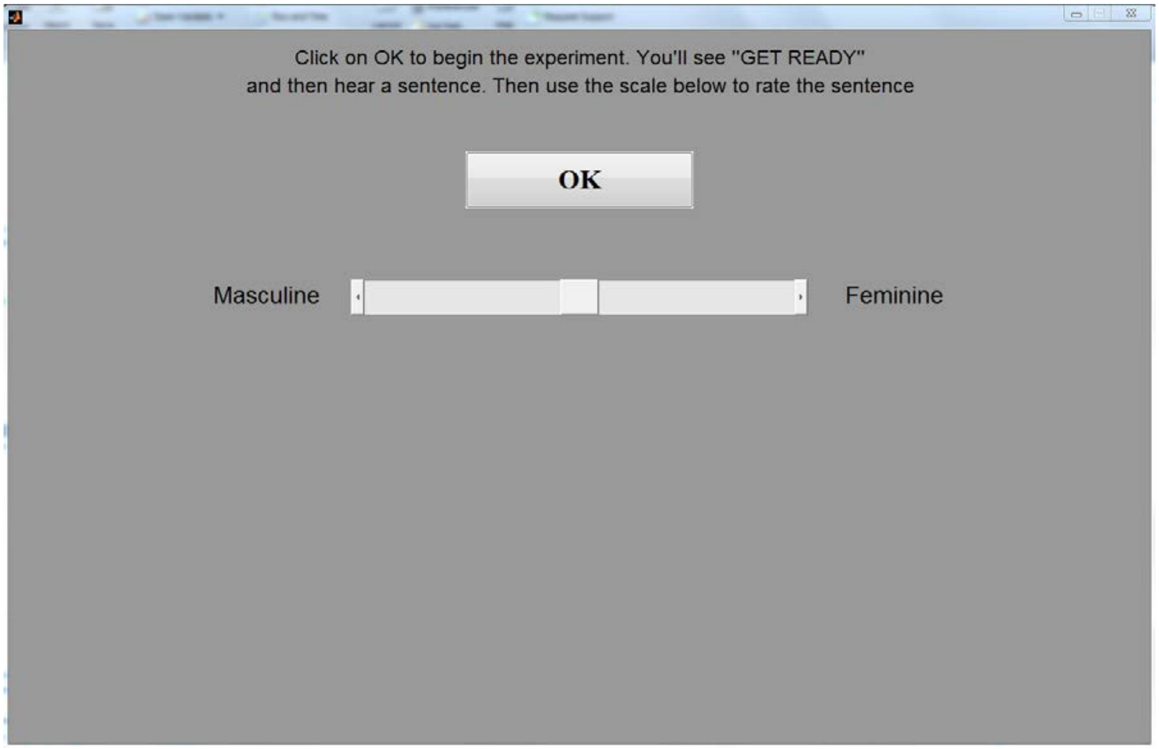

Figure 2.1. Response screen used by participants to rate each sentence.

CHAPTER 3

RESULTS

Each token was identified 17 times (once per listener), totaling 11,152 observations. Overall femininity ratings were calculated by averaging all listener ratings for all conversational tokens and clear tokens for each talker. These values are shown in Table 3.1.

The hypothesis, as stated, was that speaking style would shift perception of femininity. All analyses were performed with the aim of determining whether that effect occurred as well as what influence other factors might have on this effect. To assess the effect of speaking style and talker gender on perceived femininity, as well as to examine any interaction between these factors, listener ratings were analyzed with linear-mixed effects models. As noted by Ferguson (2012), the advantage of using mixed-effects models is "their ability to simultaneously account for multiple sources of inter-correlation" (p. 782). In this study, for instance, a single listener's femininity ratings may be correlated across talkers, or a single talker's femininity scores may be correlated across listeners. Speaking style and talker gender were fixed effects and listener, talker, and sentence number were considered random effects. All analyses were carried out using Stata 14.1 (StataCorp, 2015).

Speaking Style Effect on Perceived Femininity

The fixed effects of gender and speaking style were both significant (β = 50.37and β = 1.42, respectively, both $p < .001$). Femininity judgments were an average of 50 percentage points higher on the femininity scale for female talkers than male talkers, as expected.

Table 3.1

Mean Perceived Femininity for All Talkers

In Clear (CL) and Conversational (CO) Speaking Styles

Code	CL	CO		Code	CL	CO
F01	78.01	78.83		M01	33.09	26.84
F02	71.36	76.63		M02	20.70	18.44
F03	73.62	71.32		M03	34.56	32.45
F04	79.54	77.83		M04	29.91	26.89
F05	79.08	81.44		M05	21.64	22.34
F06	86.96	87.14		M06	26.16	20.51
F07	74.87	80.90		M07	17.29	16.40
F08	72.79	79.98		M08	11.64	6.34
F09	71.89	77.42		M09	21.02	13.21
F10	71.13	67.55		M10	25.82	11.68
F11	74.20	72.88		M11	29.96	23.52
F12	74.71	69.09		M12	30.26	24.68
F13	61.99	62.43		M13	30.32	28.81
F14	84.61	85.35		M14	25.50	20.73
F15	73.62	74.32		M15	21.48	22.81
F16	62.25	65.15		M16	27.51	27.74
F17	67.57	64.39		M17	27.65	26.40
F18	60.60	61.83		M18	30.24	28.88
F19	69.60	68.85		M19	13.12	11.32
F20	70.49	70.46		M20	19.79	14.78
F21	75.15	75.71				
MEAN	**73.05**	**73.78**		**MEAN**	**24.88**	**21.24**

Although the effect of speaking style was significant, it was extremely small (only 1.43 percentage points on average). Figure 3.1 displays average femininity ratings in both speaking styles for both gender groups.

The interaction between gender and speaking style was also significant (β = -4.33, $p < .001$), and so a stratified analysis was undertaken to determine the source of the interaction. First, the effect of gender was tested for each speaking style; the effect was large and significant for both clear speech (β = 48.21, $p < .001$) and conversational speech (β = 52.54, $p < .001$), as expected. Then the effect of speaking style was tested for each gender. While the effect of speaking style was statistically significant in both cases, the effect was larger for male talkers (β = 6.34, $p < .001$) than for female talkers (β = -0.69, $p = .025$). A line graph of all talkers, Figure 3.2, ranked by average perceived femininity ratings in conversational speech, displays this discrepancy nicely, as well as highlighting the degree of variability between individual talkers.

There was a possibility that specific sentences might influencing gender perception given that vowel space and formant frequencies are among the acoustic correlates of perceived femininity outlined in the gender perception literature. To assess this influence, the analysis was repeated with sentence number as a third random factor. Comparison of this model with the previous two-random-factor model indicated that adding sentence number as a random effect did not account for any additional variance. In other words, individual sentences did not have a significant effect on perceived femininity.

Speaking Style Effect on Acoustic Metrics

To validate the assumption that clear speech does indeed change acoustic parameters typically associated with feminine speech, mixed-effects models were performed on each

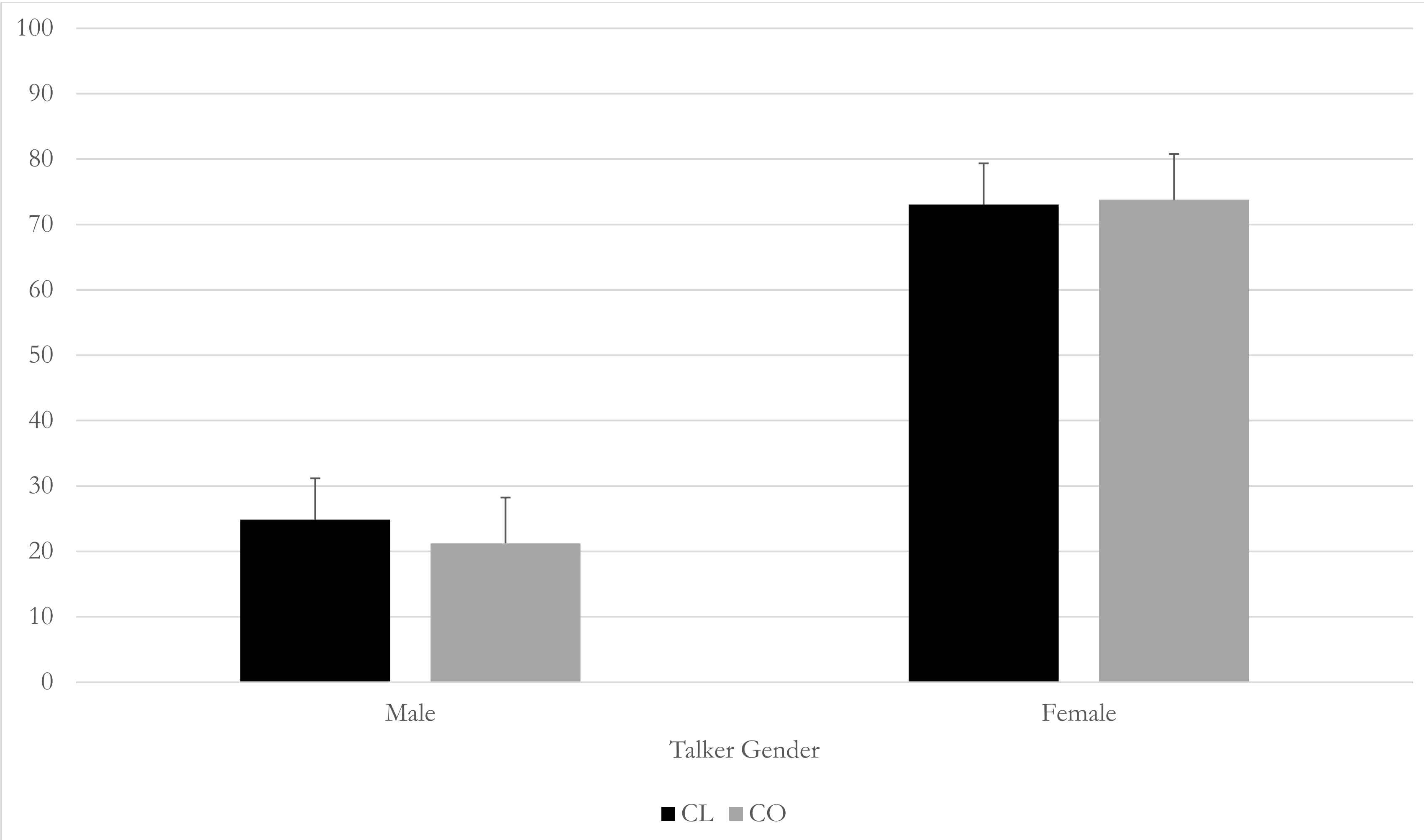

Figure 3.1. Mean femininity ratings by gender in both speaking styles.

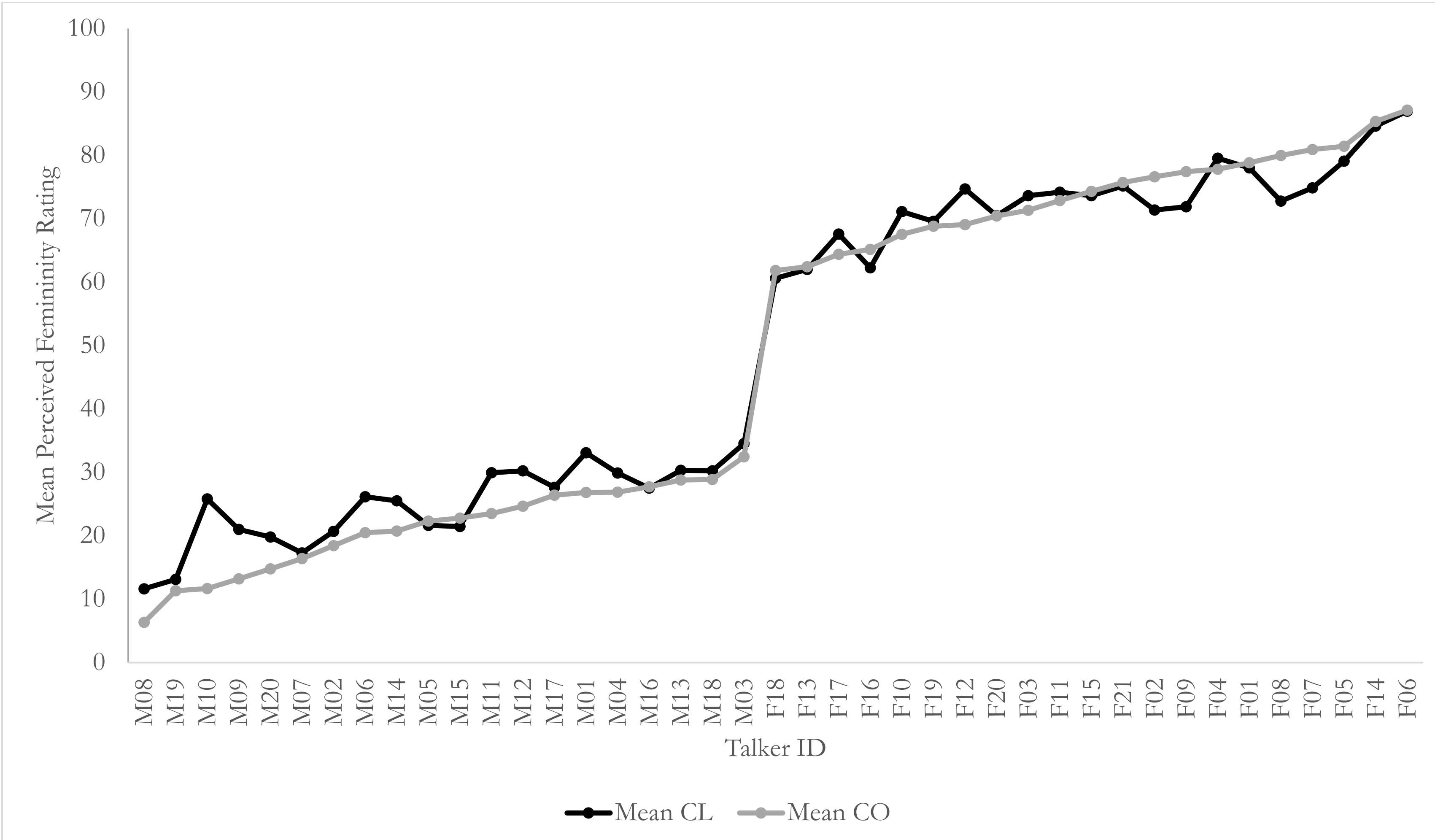

Figure 3.2. Mean femininity ratings by talker in both speaking styles. Data points are ranked by femininity ratings in conversational (CO) speaking condition.

acoustic metric with talker gender and speaking style as fixed effects and individual talker as a random effect. Acoustic metrics tested included mean f_o, mean f_o standard deviation, mean speaking rate, mean CPPS, and mean vowel perimeter.

Values for each of these metrics were obtained via custom script in Praat acoustic analysis software (Boersma & Weenik, 2015) to determine mean f_o, f_o standard deviation, speaking rate, and CPPS for each sentence. Due to the high error observed in f_o tracking, each token was reviewed by the experimenter and adjusted by hand to eliminate incorrectly tracked f_o contours. Observed errors typically occurred either in segments that contained significant glottal fry or a high amplitude fricative, which were both labeled in error as high frequency voicing. Adjustments consisted of either changing pitch parameters within Praat's pitch options, adjusting pitch contour within the pitch object itself, or both. Once adjusted, mean f_o and f_o standard deviation were recalculated for these sentences. It is important to note that glottal fry is not typically marked as f_o in acoustic analysis, so the choice to mark these segments with pitch contours is not consistent with current analysis protocols. This does have some impact on the validity of computed means. Means for each speaking style were then calculated for each of these metrics for each talker.

Speaking rate was obtained by subtracting 100 ms from the total duration of each stimulus file to account for the added silence. The resulting number was divided by eight, resulting in the syllables per second speaking rate for each sentence. These values were also averaged for both speaking styles for each talker.

Vowel perimeter is calculated by extracting a vowel's steady-state F1 and F2 values, averaging these values across two representative tokens, and then adding the Euclidian distances between these values for /i/ and / æ /, / æ / and /a/, /a/ and /u/, and /u/ and /i/ in Barks (Ferguson et al., 2010). Individual talker data for this metric were obtained via communication with the first author of that paper, as only group means were reported in publication. Note that the sentences used in the present perceptual experiments were chosen from the Ferguson Clear Speech Database with no consideration of the sentences which contained the /bVd/ tokens used by Ferguson et al. (2010) to compute vowel space perimeter.

The interaction between speaking style and talker gender was not significant for mean f_0 (β = 4.61, p = .378). The main effect of talker gender on mean f_0 was significant (β = 73.52, p < .001), as expected: female talkers had a mean f_0 73 Hz higher on average than male talkers. The speaking style effect on mean f_0 was also significant (β = 11.22, p < .001). The mean f_0 in clear speech was, on average, 11 Hz higher than in conversational speech.

The interaction between talker gender and speaking style was significant for f_0 standard deviation (β = 5.02, p = 0.027). Analysis of the main effects indicated a significant effect of gender, namely, female talkers had more variable pitch than male talkers (β = 13.43, p < .001). The main effect of speaking style was not significant overall (z = -0.28, p = 0.778). To determine the source of the interaction, a stratified analysis was performed. Analysis of the gender effect in the two speaking styles showed, in clear speech, female talkers had more f_0 variability than male talkers (β = 13.43, p < .001). Female talkers also had more variability than male talkers in conversational speech (β = 18.48, p < .001). Analysis of the speaking style effect for the two gender groups showed that male talkers had significantly more f_0 variability in clear speech (β = 2.91, p < .001). The effect of speaking style on f_0 variability was not significant for female talkers (z = 0.96, p = 0.25).

The interaction between speaking style and talker gender was not significant for mean speaking rate (β = -0.03, p = 0.244). The main effect of talker gender was significant

(β = 0.028, p = 0.02); speaking rate was slower for female talkers (M = 0.283 syllables per second) than for male talkers (M = 0.254 syllables per second). The main effect of speaking style was also significant (β =-0.095, $p < .001$), showing the expected reduction in speaking rate for clear speech. These results suggest that, on average, male and female talkers showed similar speaking rate reductions when they spoke clearly.

The interaction between speaking style and talker gender was not significant for mean CPPS (z = 0.9, p = 0.367). Interestingly, the main effect of talker gender on mean CPPS was not significant (z = -0.91, p = 0.363), which implies no difference between measures of breathiness in male or female talkers for the present test materials. The main effect of speaking style, however, was significant (z = -3.47, p = .001). Namely, CPPS is higher in clear speech, which indicates less breathiness. Eliciting clear speech from a speaker reduced the amount of breathy vocal quality that speaker produced.

The interaction between speaking style and talker gender was not significant for vowel perimeter (β = -0.34, p = 0.34). The main effect of talker gender on vowel perimeter was significant (β = 1.70, $p < .001$), indicating a bigger vowel space for female talkers. The main effect of speaking style was also significant (β = 1.13, $p < .001$), indicating a larger vowel space in clear speech. Using a clear speech strategy typically increased the vowel space of all speakers, and female speakers typically produced vowels with larger vowel perimeter overall.

Acoustic Correlates

Keeping in mind that clear speech had a significant effect on femininity, and that there is significant variability in how individual talkers adjusted their speech to speak clearly (e.g., Ferguson & Quené, 2014), it can be posited that specific acoustic changes may be associated with shifts in perceived femininity. When examining Figure 3.2, we can predict that those talkers with the biggest perceived femininity differences between the two speaking styles may have had larger acoustic differences between speaking styles. The acoustic characteristics discussed in Chapter 1, mean fundamental frequency (f_0), mean f_0 standard deviation, vowel space, CPPS, and speaking rate, were of the most interest.

Means for each of these acoustic metrics are listed in Appendix C. Pairwise correlational analyses were carried out using Stata 14.1 to examine the relationship between the mean values of each acoustic variable and mean listener ratings of femininity for each talker in each speaking style. Because the distribution of the data was bimodal, clearly separated by gender group, each gender group was correlated separately.

Correlational analyses of average f_0 and perceived femininity for both gender groups revealed a slightly stronger relationship for male talkers than female talkers ($r = 0.70$ and $r = 0.61$, respectively, both $p < .001$). The scatterplot in Figure 3.3 illustrates this relationship.

f_0 variability, measured as f_0 standard deviation, was moderately and significantly correlated with perceived femininity for female talkers ($r = 0.66$, $p < .001$), more than for male talkers ($r = 0.57$, $p < .001$). Figure 3.4 illustrates this relationship.

The speaking rate metric, measured in syllables per second, was not significantly correlated with perceived femininity within gender categories (females, $r = -0.19$, $p = 0.23$; males, $r = 0.18$, $p = 0.27$). This relationship is visualized in Figure 3.5.

For the measure of breathy vocal quality, CPPS, recall that lower CPPS numbers indicate more breathiness or roughness, as this number represents a ratio of turbulence to the voice signal. The correlation between CPPS and perceived femininity was not significant for either male talkers ($r = .04$, $p = .81$) or female talkers ($r = -0.137$, $p = 0.388$). In other words, breathiness did not influence perception of talker femininity. These relationships are

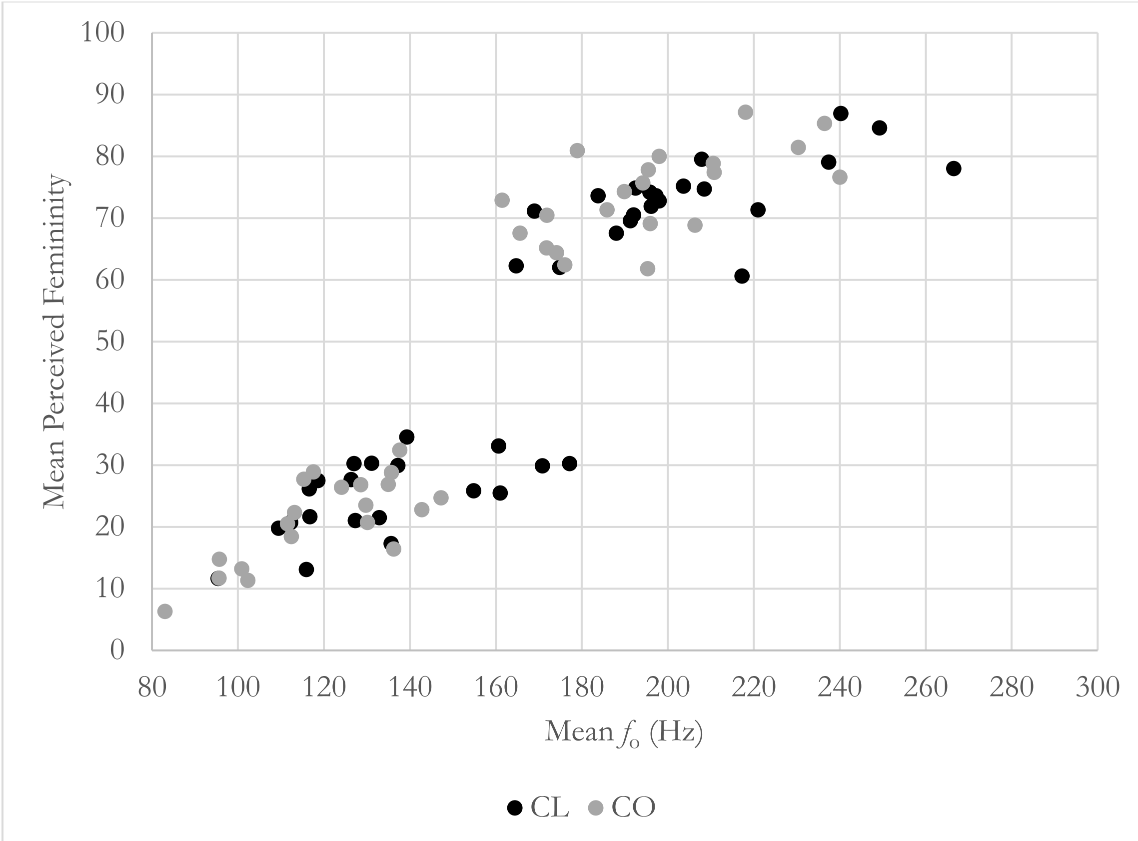

Figure 3.3. Correlations between Mean f_o and perceived femininity for all talkers in both speaking styles. The bottom left data point cloud corresponds to male talkers; the top right cloud corresponds to female talkers.

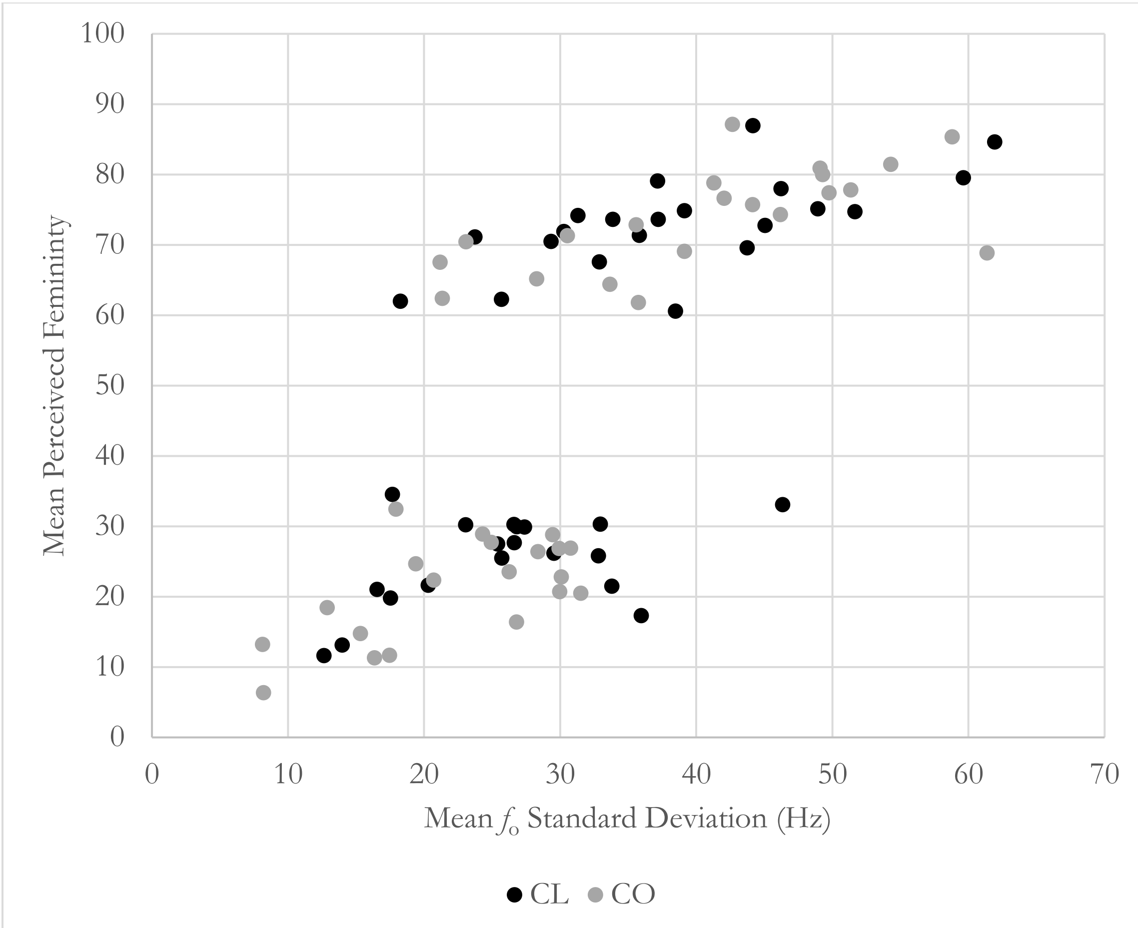

Figure 3.4. Correlations between mean f_o standard deviation and perceived femininity for all talkers in both speaking styles.

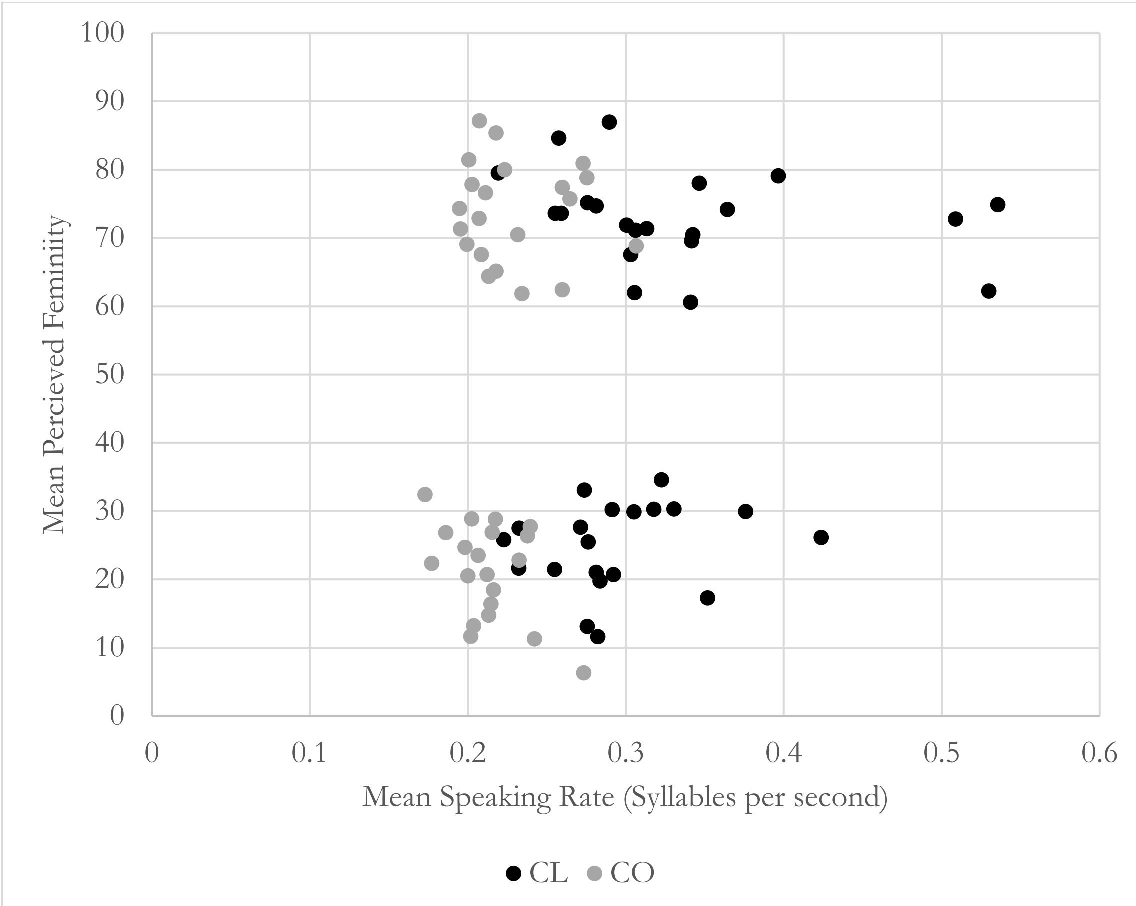

Figure 3.5. Correlations between speaking rate and perceived femininity for female all talkers in both speaking styles.

illustrated in Figure 3.6. The correlation between perceived femininity and vowel space for the two gender groups yielded interesting findings. Vowel space was correlated with perceived femininity only for male talkers (r = .49, p = .001), and not female talkers (r = -0.07, p = .67). These correlations are displayed in Figure 3.7.

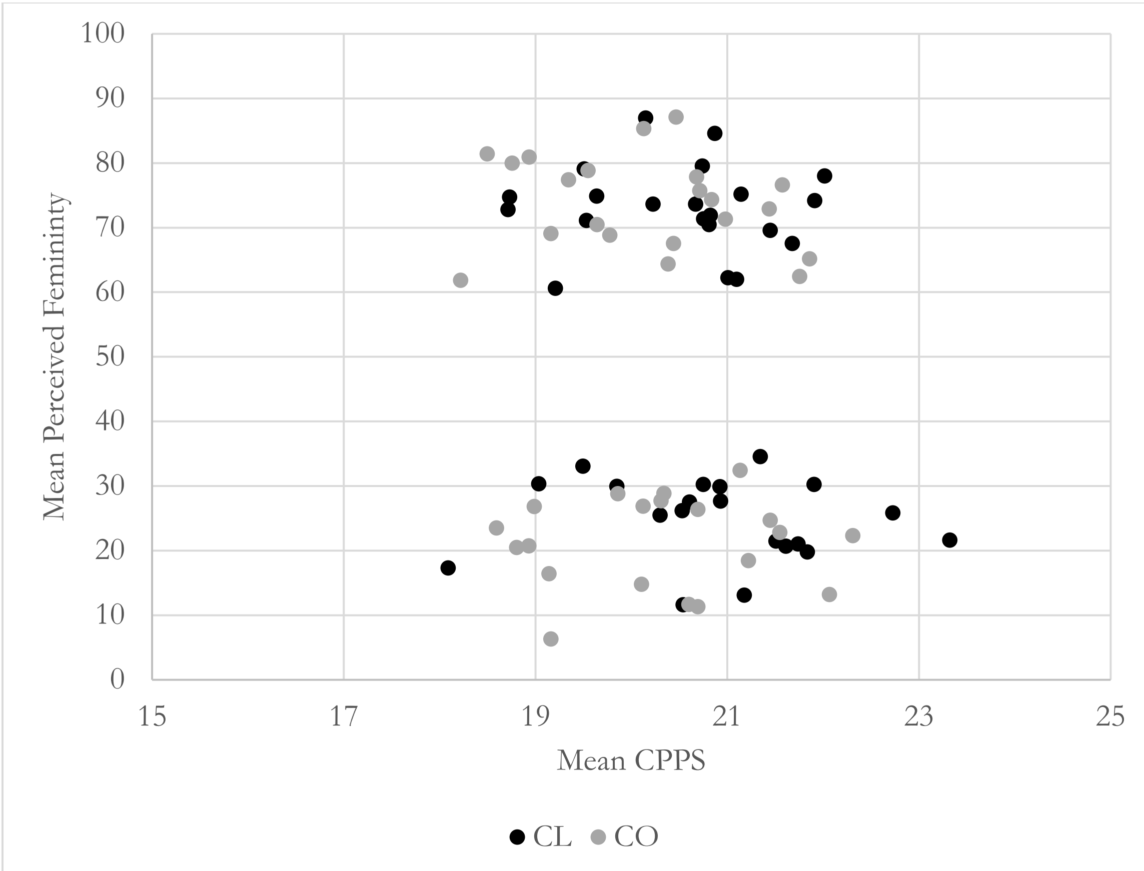

Figure 3.6. Correlations between mean Cepstral Peak Prominence (CPPS) and perceived femininity for all talkers in both speaking styles.

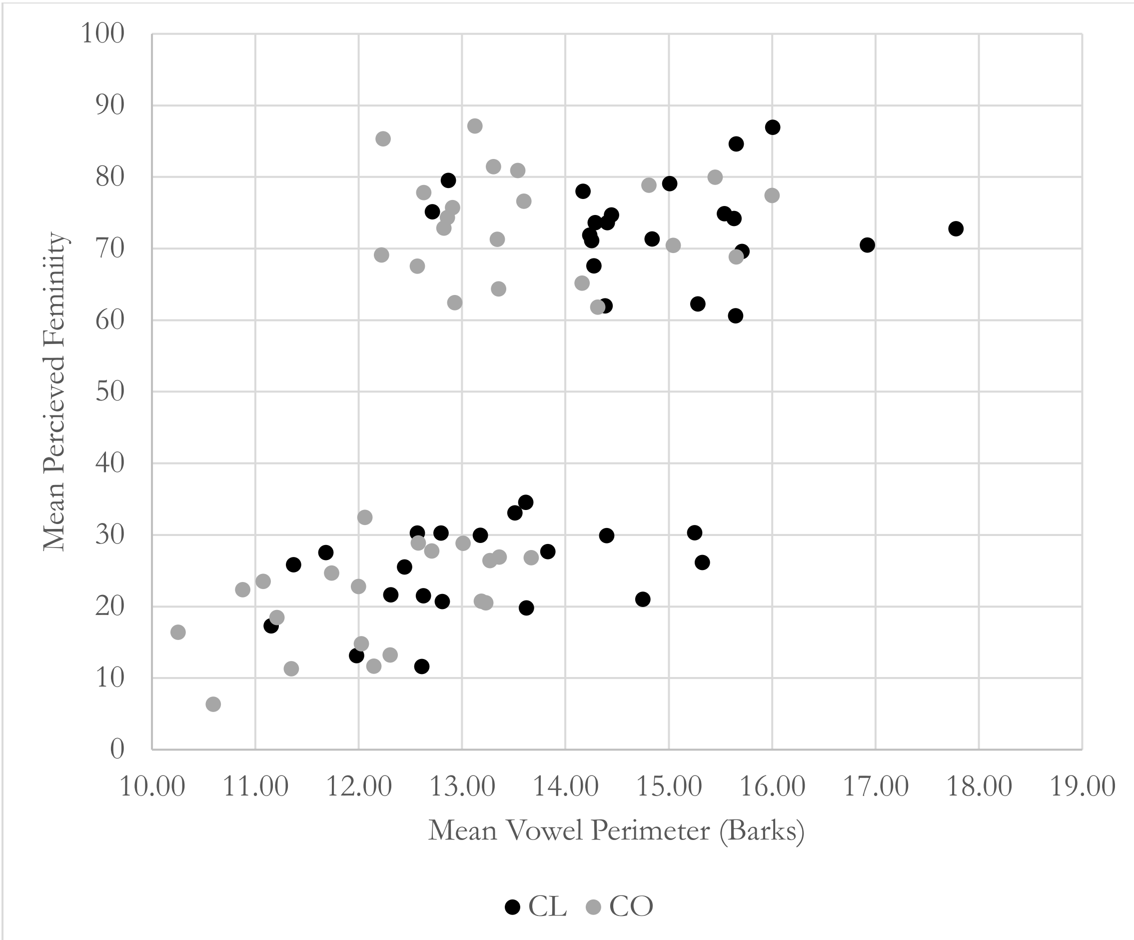

Figure 3.7. Correlations between mean vowel perimeter and perceived femininity for all talkers in both speaking styles.

CHAPTER 4

CONCLUSION AND DISCUSSION

Significance of Findings

In many ways, the findings of the present study validated what is already known about gender perceived through voice. Higher f_o, greater f_o variability, and more expanded vowel space all contributed to greater perceived femininity, as outlined in the gender perception literature. Given that adopting a clear speaking style also influenced these acoustic changes, it is no surprise that the initial hypothesis was supported: clear speech increased listeners' perception of femininity. However, the average effect of adopting a clear speaking style was small, and because of the large degree of talker variability, not all talkers enjoyed this effect, despite receiving identical instructions for producing clear speech. The interesting outcomes of this study, then, lie in the acoustic details that separate one talker's clear speech from another's.

While this study examined changes in gender perception for female as well as male talkers, the clear speech effect on perceived femininity for male talkers is of most interest if it is to be applied to transgender female speakers. It is no surprise that shifting f_o upward changes listeners' perception of femininity for this talker group. While talkers produce clear speech that is, on average, higher f_o than conversational speech, change and degree of change in f_o were highly variable among talkers. For example, talker M10 produced clear speech with an average pitch 59 Hz higher than conversational speech, on average, and was rated 14 percentage points more feminine as a result. Talker M09 produced clear speech 26 Hz higher, and was rated 8 percentage points more feminine. Although male talkers who increased f_o changed perceived femininity ratings, the same was not true for female talkers. For instance, talker F01 produced clear speech 56 Hz higher, on average, but received nearly identical femininity ratings in both speaking styles. Talker F06 experienced similar results with a 22 Hz difference. It is likely that a talker's optimal f_o, the pitch that is determined largely by a talker's vocal mechanism, plays the largest part in perceptions of femininity. Talker M10 may have achieved a large change in femininity ratings, but he was rated one of the lowest on the femininity scale overall. Conversely, talker F01 was rated as one of the highest. There may be a limit to how feminine a voice can be perceived once f_o passes above a certain frequency. Similarly, changes in f_o for voices at the lower extremes of human vocal range may achieve more change in femininity ratings than those with higher ranges, which would explain why f_o was somewhat more strongly correlated with perceived femininity in male talkers than female talkers.

Given the multidimensional nature of the human voice, it is unsurprising that no one parameter may be lauded as solely responsible for changes in listener perception of gender. Despite the primacy of f_o as a cue of a talker's gender, the other acoustic metrics examined in this study paint a more nuanced picture of gender differences in clear speech, as well as how those differences are perceived by listeners. Female talkers' f_o variability across speaking styles was greater than male talkers', consistent with previous studies' findings about behavioral gender differences in speech prosody. Using more exaggerated prosody resulted in higher perception of femininity, overall, and male talkers used greater f_o variability in clear speech than in conversational speech. In other words, eliciting a clear speaking style results in greater f_o variability for male talkers, who have been socialized not to use as much f_o

variability, which in turn increased perception of femininity for that talker group. For a talker attempting to tip the femininity scales, exaggerating prosody might help achieve that. Reviewing the specific talkers above, talker M10 produced a 15 Hz f_o standard deviation increase in clear speech, which likely contributed to his large change in perceived femininity. Talker M09, similarly, produced an 8 Hz f_o standard deviation increase. The female talkers above who did not shift femininity ratings, talker F01 and F06, produced similar f_o variation in both speaking styles, which may explain why their femininity ratings did not change between styles. However, these talkers used more prosodic variation than the talkers M09 and M10, a possible contributor to their high overall femininity ratings.

Findings regarding vowel perimeter were particularly interesting in light of previous perceived gender literature. Higher formant frequencies and increased vowel space have been correlated with perceived femininity in other studies, and resonance training has been successful for transgender women attempting to sound more feminine (Carew, Dacackis, & Oates, 2006). Female talkers in this study did not significantly change their femininity ratings by using clear speech, likely because they produced conversational speech that already had a greater vowel space than male talkers. Male talkers, despite not attempting to sound "feminine," expanded their vowel space and were perceived as more feminine in clear speech. Eliciting a clear speaking style encourages greater excursion of the speech structures for both male and female talkers, but resulted in increased perceived femininity only for the male talkers. Differences in correlation between the two groups was much greater for measures of resonance than for measures of pitch, pointing to the necessity of including resonance training for talkers wishing to shift perceived femininity.

The other two acoustic metrics examined in this study, speaking rate and CPPS, did not affect listener judgments of femininity. Given that slower speech rate was noted as a feminine voice characteristic in the gender perception literature, and that clear speech is consistently longer in duration than conversational speech, it is surprising that there was no relationship between speaking rate and perceived femininity. It is possible that a speaking rate measure, such as syllables per second, does not give a clear indication of *how* talkers achieve a slower rate. It is possible to lengthen an utterance through increasing the syllable length or through increasing pause lengths. Clinicians who provide voice feminization training teach their clients to use a "legato" speaking style, that is, to increase their vowel length and decrease their pause length (Adler, Hirsch & Mordaunt, 2012). Examining these details may explain how speaking rate is related to perceived femininity; however, it is also possible that speaking rate is not salient at the sentence level in shifting gender perception. It may be necessary to examine speaking rate at the narrative level to further clarify this relationship.

The second hypothesis regarding breathy vocal quality was supported: there was no correlation between measures of breathiness and perceived femininity. CPPS had not yet been measured for the Ferguson Clear Speech Database, but the increase in vocal effort and increased intensity that accompanies clear speech seemed likely to reduce breathy vocal quality. Results showed this to be true: conversational speech was more breathy than clear speech. Van Borsel et al. (2009), which demonstrated a correlation between breathy vocal quality and perceived femininity, had a completely different methodology than the present study. Talkers in Van Borsel et al. were trained to produce more breathy vocal quality, and CPPS measures for the resulting stimuli used would likely show significantly lower CPPS values (i.e., more breathiness) for the trained items than the untrained tokens. As the talkers in the Ferguson Clear Speech Database did not attempt to alter their breathiness, and that all talkers had normal voice quality, it is likely that any differences in breathiness were not perceptible to participants. While training a speaker to use breathier vocal quality may be appropriate in some scenarios, it isn't clear what impact it has on perceived femininity for transgender speakers.

In all, for the three metrics most highly correlated with perceived femininity, average f_0, f_0 variability, and vowel space, eliciting clear speech from male talkers created larger changes in these metrics than for female talkers, and female talkers exhibited these characteristics across speaking styles. In other words, cueing someone with a male vocal mechanism and socialized vocal behavioral patterns to use clear speech is likely to increase their perceived femininity. Is this because female talkers were already perceived as feminine, and thus were less likely to shift their femininity, regardless of the changes they made? Future investigations might attempt to determine whether the parameters with which listeners judge the male voice are narrower than those used to judge female voices.

Limitations

As with all speech elicited in laboratory conditions, the speech materials used in the present study are not entirely representative of what speakers use in their daily lives, and as such, may not represent all of the changes speakers make when communicating with a partner who is having difficulty understanding them. Further, because the speech materials were read, the "conversational" speech cannot be truly conversational. These limitations reflect necessary controls put in place to limit semantic content, ensure all vowels can be examined in the same phonemic context, and tightly regulate the sound conditions under which speech is recorded, such that analysis of these recordings can yield useful information (Xu, 2010). In doing so, talkers may limit their prosodic variation, and the neutral sentences they read are largely void of emotional content, which may result in a flattened vocal affect. f_0 range, variation, and prosodic contour shape were particularly important acoustic correlates of perceived femininity in the gender perception literature, so a study of laboratory speech may not capture the changes talkers make in their entirety. It is important to note, however, that clear speaking style has been examined in spontaneous speech with similar results (Hazan & Baker, 2011), though spontaneous speech recordings were not used for this study.

Some listener effects may exist in this study given the mean age and regional origin of participants. There are regional and generational differences in attitudes toward gender that may not be captured in a study where participants mean age was not much greater than 20 and nearly all of them had resided in a socially conservative state for most of their lives. Including participants from more than one generational group and participants from a wider range of cultures and regions may produce different results.

The fact that speakers in this study were not attempting to change their presented vocal femininity limits application of the results to some extent. We cannot extrapolate what might result for transgender speakers using clear speech from a set of recordings that only included cisgender speakers. Beyond talker characteristics, neutral sentences do not include lexical, pragmatic, and nonverbal differences that transgender speakers may or may not use, so interpretation of the results of this study are limited to *vocal* femininity only.

Directions for Future Work

One pervasive question that arises while considering the perceived gender literature is this: with which underlying construct do listeners rate the voices they hear? Studies that offer two choices, male and female, suggest to listeners that they must choose the sex of the speaker, and fail to capture gender ambiguity or degrees of "maleness" or "femaleness." We know that some male voices are perceived as more feminine, and that conscious adjustments can be made that impact perception, even when the voice is not perceived as crossing the gender boundary. The social impact of gender presentation via voice for those voices perceived as nonconforming is generally negative, as is the case with "gay" speech, some voice pathologies, and some transgender voices. This highlights the importance of understanding not only which adjustments and which acoustic parameters influence the listener's perceptions of gender, but also *how much*. For those talkers who want to tip the scales with a variety of changes, even small changes become impactful.

Providing listeners with a visual analog scale allowed such small changes in perception to be captured in this study; however, it is possible that participants still used binary gender categories to make their ratings. When presented with a scale from 0 to 100, listeners regularly split the scale, rarely rating a female voice below 50 or a male voice above 50. It is entirely likely that a participant heard the sentence, determined the speaker's sex, and rated the voice using the corresponding half of the scale. Thus, a very "feminine" sounding male voice would still never approach 100. This prompts the question: what basis for comparison did listeners use? A study that presented only male voices, but still used a scale from masculine to feminine, might have completely different outcomes as listeners would be making their femininity ratings based on only male voices, rather than comparing them with female voices. Regardless of what construct listeners use to make their judgments, or with which scales they are presented, it is clear from the literature that gender perception is a fairly automatic and robust listener judgment. As such, influencing listener's ratings by altering the presentation or the response method may have little impact compared to behavioral changes in the speaker, such as raising f_0.

Regardless of what underlying construct a listener uses, would it be possible to reset that construct? After all, this study and many others that examined perceived femininity operated under the assumption that *speakers* must make some or all the changes necessary in order to impact listener perceptions. A perceived gender study that gives baselines for comparison that expand the vocal gender boundaries, for example, a male voice meant to be rated somewhere above 50 on the femininity scale, and assesses changes in listener's willingness to judge a "male" voice as feminine could provide interesting information about how malleable the vocal gender categories are. Ideally, some of the burden of reducing negative judgments of voices that fall outside of the typically accepted gender categories should lie with the listener. A study that supports the ability of listeners to change their vocal gender constructs would be valuable in advocating for those who are subject to these negative judgments.

In the interim, this study would be more meaningful in its application to transgender speakers if transgender speakers produced the stimuli. Recruiting transgender women who would like to impact listener's perceptions of their femininity and having them produce clear and conversational speech would provide some validation that the effect of speaking style translates to the intended population.

Further analysis of the data obtained in this study are also needed. A step-wise regression analysis of the examined acoustic variables would further elucidate the impact of each variable and provide more insight into their predictive value on ratings of perceived femininity. In addition, several male talkers' acoustic profiles put them solidly in the gender-ambiguous range, despite being rated well below 40 on the femininity scale, on average. Further analysis of specific talkers would similarly provide information about relative contributions of each acoustic variable to listener ratings of femininity.

Clinical Applications

Ideally, transgender voice and communication training encompasses many facets of communication, including pitch, prosody, resonance, semantics, pragmatics, and nonverbal communication. This study illuminates one possible set of cues that could be used in attempting to address at least three of these factors: pitch, prosody and resonance. Clear speech is quick and easy to elicit, and if it produces the desired change, could be a very powerful cue for clinicians or self-cue for clients.

This application is limited by the large degree of variability in how individual talkers produce clear speech. Simply telling a client to speak as though they are talking to someone who cannot understand them may not produce the desired result. Clear speech may or may not result in globally higher f_o, for instance. For a transgender woman with a particularly low optimal f_o, speaking clearly may not shift listener perceptions of her voice enough for her comfort.

The other two vocal parameters, prosody and resonance, are more impacted by adopting a clear speaking style, and as f_o is not the only acoustic correlate of feminine speech, having tools available for speakers to adjust these variables simultaneously could be of some use. A clear speech strategy seems particularly well-suited to eliciting greater vowel space, and could be used in combination with other strategies that address f_o, prosody, and so forth. However, adopting a clear speaking style, for some speakers, may result in changes in speech naturalness, such as abnormally slowed speech rate or over-articulated consonants. At best, clear speech is one tool available to clients and clinicians, but because additional coaching would be required, it has limited clinical utility.

APPENDIX A

LIST OF SENTENCE STIMULI

1. Use the word bad in a sentence.

2. Use the word bod in a sentence.

3. Use the word bode in a sentence.

4. Use the word bud in a sentence.

5. They spelled the word bade the wrong way.

6. They spelled the word bed the wrong way.

7. They spelled the word bood the wrong way.

8. Write the word bead on the chalkboard.

9. Write the word bode on the chalkboard.

10. Write the word bude on the chalkboard.

APPENDIX B

PARTICIPANT INFORMATION SHEET

Speech Perception Laboratory

Participant Information Sheet

Subject ID: Age: Gender: Today's date:

1. Do you have any history of speech, language, or hearing disorders?

   If so, please describe.

2. Are you a native speaker of American English?

   If not, what is your first language or dialect of English?

3. Where did you grow up?

4. How long have you lived in the Salt Lake City area?

5. Do you talk like other people who live here, or do you have an accent?

   If you have an accent, please describe.

6. It is important that the ethnic and racial makeup of our research participant pool reflects that of the local community. Please indicate which of the following ethnic and racial categories you identify with by checking the box next to the category:

Ethnic Category	
Hispanic or Latino	
Not Hispanic or Latino	
Prefer not to identify	

Racial Categories	
American Indian/Alaska Native	
Asian	
Native Hawaiian or other Pacific Islander	
Black or African American	
White	
Prefer not to identify	

APPENDIX C

ACOUSTIC DATA FOR ALL TALKERS

Table C.1

Summary of Acoustic Data for All Talkers in Clear (CL) Speaking Style

Code	CL Mean f_o	CL f_o Standard Deviation	CL Mean Syllables per Second	CL Mean CPPS	CL Mean Vowel Perimeter
F01	266.51	46.23	0.35	22.02	14.17
F02	220.96	35.82	0.31	20.75	14.84
F03	183.73	33.85	0.26	20.23	14.41
F04	207.89	59.60	0.22	20.74	12.87
F05	237.41	37.16	0.40	19.51	15.01
F06	240.25	44.15	0.29	20.15	16.01
F07	192.49	39.12	0.54	19.64	15.54
F08	198.01	45.05	0.51	18.71	17.78
F09	196.09	30.25	0.30	20.83	14.24
F10	168.94	23.72	0.31	19.53	14.26
F11	195.85	31.28	0.36	21.91	15.63
F12	208.49	51.64	0.28	18.73	14.45
F13	174.80	18.25	0.31	21.10	14.39
F14	249.24	61.91	0.26	20.87	15.65
F15	197.22	37.18	0.26	20.67	14.29
F16	164.73	25.68	0.53	21.01	15.28
F17	187.99	32.88	0.30	21.68	14.28
F18	217.24	38.46	0.34	19.21	15.65
F19	191.30	43.73	0.34	21.45	15.71
F20	192.02	29.32	0.34	20.81	16.92
F21	203.62	48.92	0.28	21.14	12.71
M01	160.61	46.34	0.27	19.49	13.51
M02	112.23	15.19	0.29	21.61	12.81
M03	139.28	17.67	0.32	21.34	13.62
M04	170.79	27.36	0.31	20.92	14.40
M05	116.71	20.29	0.23	23.32	12.31
M06	116.60	29.52	0.42	20.53	15.33
M07	135.61	35.95	0.35	18.09	11.15
M08	95.36	12.62	0.28	20.54	12.61

Table C.1 Continued

Code	CL Mean f_o	CL f_o Standard Deviation	CL Mean Syllables per Second	CL Mean CPPS	CL Mean Vowel Perimeter
M09	127.32	16.54	0.28	21.74	14.75
M10	154.83	32.79	0.22	22.73	11.37
M11	137.22	26.80	0.38	19.85	13.18
M12	177.16	26.59	0.32	21.90	12.80
M13	131.10	32.94	0.33	19.03	15.25
M14	161.02	25.71	0.28	20.30	12.44
M15	132.87	33.80	0.25	21.51	12.63
M16	118.59	25.39	0.23	20.61	11.68
M17	126.34	26.62	0.27	20.93	13.83
M18	126.97	23.05	0.29	20.75	12.57
M19	115.87	13.97	0.28	21.18	11.98
M20	109.44	17.53	0.28	21.83	13.62

Table C.2

Summary of Acoustic Data for All Talkers in Conversation (CO) Speaking Style

Code	CO Mean f_o	CO f_o Standard Deviation	CO Mean Syllables per Second	CO Mean CPPS	CO Mean Vowel Perimeter
F01	210.52	41.28	0.28	19.55	14.81
F02	239.99	42.03	0.21	21.58	13.60
F03	185.85	30.51	0.20	20.98	13.34
F04	195.43	51.35	0.20	20.68	12.63
F05	230.34	54.28	0.20	18.49	13.30
F06	218.07	42.65	0.21	20.47	13.13
F07	178.90	49.09	0.27	18.93	13.54
F08	197.95	49.27	0.22	18.75	15.45
F09	210.74	49.74	0.26	19.34	16.00
F10	165.62	21.15	0.21	20.44	12.57
F11	161.41	35.58	0.21	21.44	12.83
F12	195.90	39.13	0.20	19.16	12.22
F13	176.01	21.34	0.26	21.76	12.93
F14	236.46	58.80	0.22	20.13	12.24
F15	189.83	46.16	0.19	20.84	12.86
F16	171.78	28.25	0.22	21.86	14.16
F17	174.06	33.63	0.21	20.38	13.36
F18	195.26	35.74	0.23	18.22	14.31
F19	206.30	61.36	0.31	19.78	15.65
F20	171.83	23.07	0.23	19.64	15.05
F21	194.19	44.11	0.26	20.71	12.91
M01	128.57	29.90	0.19	18.99	13.67
M02	112.41	12.86	0.22	21.22	11.21
M03	137.60	17.93	0.17	21.13	12.06
M04	134.93	30.75	0.22	20.12	13.36
M05	113.13	20.68	0.18	22.31	10.88
M06	111.54	31.50	0.20	18.80	13.23
M07	136.25	26.77	0.21	19.14	10.25
M08	83.00	8.18	0.27	19.16	10.60
M09	100.84	8.13	0.20	22.06	12.31
M10	95.61	17.45	0.20	20.60	12.15
M11	129.73	26.24	0.21	18.59	11.08
M12	147.24	19.37	0.20	21.45	11.74
M13	135.72	29.44	0.22	19.86	13.01

Table C.2 Continued

Code	CO Mean f_o	CO f_o Standard Deviation	CO Mean Syllables per Second	CO Mean CPPS	CO Mean Vowel Perimeter
M14	130.10	29.95	0.21	18.93	13.19
M15	142.80	30.08	0.23	21.55	12.00
M16	115.30	24.93	0.24	20.31	12.71
M17	124.12	28.35	0.24	20.69	13.27
M18	117.52	24.29	0.20	20.34	12.58
M19	102.28	16.35	0.24	20.69	11.35
M20	95.68	15.32	0.21	20.11	12.03

REFERENCES

Adler, R. K., Hirsch, S., & Mordaunt, M. (2012). *Voice and communication therapy for the transgender/transsexual client: A comprehensive clinical guide*. San Diego, CA: Plural Publishing.

Avery, J. D., & Liss, J. M. (1996). Acoustic characteristics of less-masculine-sounding male speech. *Journal of the Acoustical Society of America, 99*(6), 3738-3748.

Awan, S. N., Giovinco, A., & Owens, J. (2012). Effects of vocal intensity and vowel type on cepstral analysis of voice. *Journal of Voice*, *26*, 670-615.

Bachorowski, J.A., Owren, M.J. (1989). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America, 106*, 1054-1063.

Boersma, P., Weenink, D. (2015). Praat: doing phonetics by computer (Version 6.0.17) [Computer program]., Retrieved from http://www.praat.org/

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech Language Hearing Research, 46,* 80-97.

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*, 591–603.

Davis, H., & Silverman, S. R. (1978). *Hearing and Deafness* (4th ed.). New York: Holt, Rinehart and Winston.

Ferguson, S. H. (2002). *Vowels in clear and conversational speech: Talker differences in acoustic features and intelligibility for normal-hearing listeners* (Unpublished doctoral dissertation). Indiana University, Bloomington.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, *116*, 2365-2373.

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, *112*, 259-271.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, *50*, 1241-1255.

Ferguson, S. H., Poore, M. A., & Shrivastav, R. (2010). Acoustic correlates of reported clear speech strategies. *Journal of Academic Rehabilitative Audiology XLIII,* 45-64.

Ferguson, S. H., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners). *Journal of the Acoustical Society of America*, *135*, 3570-3584.

Ferguson, S.H., Morgan, S.D., Rogers, L.R., & Hunter, E.J. (2014, September). Within-session stability of acoustic features of conversational and clear speech for male and female talkers. *InfoFair 2014: Women's Health, Sex & Gender Research Conference*, University of Utah Health Sciences Center.

Freidenberg, C. B. (2002). Working with male-to-female transgendered clients: Clinical considerations. *Contemporary Issues in Communication Science and Disorders*, *29*, 43-58.

Garrett, R. (2013). Cepstral-and spectral-based acoustic measures of normal voices. (Unpublished master's thesis). University of Wisconsin, Milwaukee.

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*, 22-33.

Goldstein, U. G. (1980). *An articulatory model for the vocal tracts of growing children* (Doctoral dissertation), Massachusetts Institute of Technology, Cambridge, MA.

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *Journal of the Acoustical Society of America, 130,* 2139-2152.

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics*, *71*, 1150-1166.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical society of America*, *97*, 3099-3111.

Hirano, M., Kurita, S., & Sakaguchi, S. (1989). Ageing of the vibratory tissue of human vocal folds. *Acta oto-laryngologica*, *107*, 428-433.

Holmberg, E. B., Hillman, R. E., Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. *Journal of the Acoustical Society of America*, *84*, 511-529.

Honorof, D. N., & Whalen, D. H. (2010). Identification of speaker sex from one vowel across a range of fundamental frequencies. *Journal of the Acoustical Society of America*, *128*, 3095-3104.

Hubbard, D. J., & Assmann, P. F. (2013). Perceptual adaptation to gender and expressive properties in speech: the role of fundamental frequency. *Journal of the Acoustical Society of America, 133*(4), 2367-2376. doi: 10.1121/1.4792145

Johnston, D. (2000). Cool Edit 2000 [Computer Program]. Phoenix, AZ: Syntrillium Software Corporation.

Kahane, J. C. (1982). Growth of the human prepubertal and pubertal larynx. *Journal of Speech, Language, and Hearing Research, 25*, 446-455.

Kazarian, A. G., Sarkisian, L. S., & Isaakian, D. G. (1978). Length of the human vocal cords by age. *Zhurnal eksperimental'noĭ i klinicheskoĭ meditsiny, 18*, 105.

Ko, S. J., Judd, C. M., & Blair, I. V. (2006). What the voice reveals: Within-and between-category stereotyping on the basis of voice. *Personality and Social Psychology Bulletin*, *32*(6), 806-819.

Lowell, S. Y., Colton, R. H., Kelley, R. T., & Mizia, S. A. (2013). Predictive value and discriminant capacity of cepstral-and spectral-based measures during continuous speech. *Journal of Voice*, *27*(4), 393-400.

Maurer, D., Landis, T. (1996). Intelligibility and spectral differences in high-pitched vowels. *Folia Phoniatr*, *48*, 1-10.

Maniwa, K., Jongman, A, Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America*, *125,* 3962-3973.

The MathWorks, Inc. (2014a). *MATLAB*. Natick, MA: The MathWorks, Inc.

Oates, J., & Dacakis, G. (1997). Voice change in transsexuals. *Venereology*, *10*, 178.

Owren, M. J., Berkowitz, M., & Bachorowski, J. A. (2007). Listeners judge talker sex more efficiently from male than from female vowels. *Perception & Psychophysics*, *69*, 930-941.

Perry, T. L., Ohde, R. N., & Ashmead, D. H. (2001). The acoustic bases for gender identification from children's voices. *Journal of the Acoustical Society of America*, *109*, 2988-2998.

Peterson, G.E., Barney, H.L. (1952). Control methods used in a study of the identification of vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29,* 434-446.

Schwartz, M. F., & Rine, H. E. (1968). Identification of speaker sex from isolated, whispered vowels. *Journal of the Acoustical Society of America*, *44*, 1736-1737.

Sodersten, M., & Lindestad, P. A. (1990). Glottal closure and perceived breathiness during phonation in normally speaking subjects. *Journal of Speech, Language, and Hearing Research*, *33*, 601-611.

StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, *85*, 1699-1707.

Titze, I. R. (1994). *Principles of voice production.* Englewood Cliffs, NJ: Prentice Hall.

Tillman, T., & Carhart, R. (1966). *An Expanded Test for Speech Discrimination Utilizing CNC Monosyllabic Words, Northwestern University Auditory Test No 6*. Brooks AFB, TX: USAF School of Aerospace Medicine.

Van Borsel, J., Janssens, J., & De Bodt, M. (2009). Breathiness as a feminine voice characteristic: A perceptual approach. *Journal of Voice*, *23*, 291-294.

Wolfe, V. I., Ratusnik, D. L., Smith, F. H., & Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorders*, *55*, 43-50.

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics, 38*, 329-336.