The Effect of Formant Biofeedback on the Feminization of Voice in Transgender Women

Deanna Kawitzky\(^1\) and Tara McAllister\(^a\)

\(^a\)New York University, New York, NY

Address correspondence to:
Tara McAllister
Department of Communicative Sciences and Disorders, New York University
665 Broadway, Suite 900
New York, NY 10012, USA
Phone: 212-992-9445
Fax: 212-995-4356
E-mail: tkm214@nyu.edu

Declarations of interest: none.

\(^1\)Present address: Burke Rehabilitation Hospital, White Plains, NY
ABSTRACT

Differences in formant frequencies between men and women contribute to the perception of voices as masculine or feminine. This study investigated whether visual-acoustic biofeedback can be used to help transgender women achieve formant targets typical of cisgender women, and whether such a shift influences the perceived femininity of speech. Transgender women and a comparison group of cisgender males were trained to produce vowels in a word context while also attempting to make a visual representation of their second formant (F2) line up with a target that was shifted up relative to their baseline F2 (feminized target) or an unshifted or shifted-down target (control conditions). Despite the short-term nature of the training, both groups showed significant differences in F2 frequency in shifted-up, shifted-down, and unshifted conditions. Gender typicality ratings from blinded listeners indicated that higher F2 values were associated with an increase in the perceived femininity of speech. Consistent with previous literature, we found that fundamental frequency and F2 make a joint contribution to the perception of gender. The results suggest that biofeedback might be a useful tool in voice modification therapy for transgender women; however, larger studies and information about generalization will be essential before strong conclusions can be drawn.

KEYWORDS: transgender, biofeedback, formants, perception, voice, fundamental frequency
1. INTRODUCTION

For individuals making the transition in presenting their gender from male to female, adjusting vocal quality to reflect this transition is often a challenging part of the process. In seeking a voice that feels authentic to the speaker and that will be perceived as congruent with a female identity, transgender women may pursue treatment from speech-language pathologists (SLPs) as specialists in the areas of vocal production, voice quality, and vocal health. In the past two decades, SLPs have been using a variety of methods to help transgender women achieve a perceptually feminine voice. Current transgender voice and communication treatment methods predominantly target speaking fundamental frequency (SFF) and resonance [1]. However, due to the dearth of research literature on transgender voice modification, it remains unclear how effectively and to what extent these methods are being implemented.

It is generally agreed that the most salient acoustic indicator of gender is fundamental frequency (F0), or pitch, its psychoacoustic correlate [2]. F0, the frequency at which vocal folds vibrate in Hertz (Hz), is partly under the active muscular control of the speaker and partly influenced by anatomical differences in vocal fold length. Because adult cisgender men tend to have a larger effective vocal fold length than adult cisgender women, their average SFF, ranging from 100-146 Hz, is lower than that of cisgender women, which averages 188-221 Hz [3]. Furthermore, research has shown that perception of gender in voice can be successfully manipulated via the alteration of F0. Specifically, adult cisgender male voices tend to be perceived as cisgender female when the F0 is artificially increased, an effect that has been replicated across the contexts of spontaneous conversation [4], scripted sentences [5], and isolated vowels [6].
Similar to F0, it is known that the acoustics of formant frequencies differ between cisgender males and females in naturally produced speech, and that this is partly attributable to differences in vocal tract shape and size. The vocal tract acts as a resonating cavity, with larger vocal tracts resonating at lower frequencies and smaller vocal tracts resonating at higher frequencies. Thus, on average, cisgender male speakers’ formants are significantly lower than cisgender female formants [7, 8]. While formant frequencies are strongly influenced by anatomical differences in vocal tract size, they are also subject to voluntary control. Formants differ within a speaker as a property of the vowel being produced and can also vary across a given speaker’s productions of a single vowel. For example, exaggerated lip rounding on /u/ will result in a lower formant frequency caused by the voluntary lengthening of the vocal tract. Given the acoustic differences between cisgender male and female formants, formants are considered an important acoustic factor contributing to the perception of gender as an indexical property of speech [9-12]. The second formant (F2), which is influenced by the length of the oral cavity, has been shown to correlate most saliently with the perception of gender [13] and was therefore chosen as the acoustic focus of the present study.

The differences between cisgender male and female formant frequencies for different vowels were quantified in a study by Fant [14]. With data from the first three formants (F1, F2, and F3) for 19 vowels across eight different languages, Fant derived proportional values or “scale factors” comparing average cisgender male and female vowel formant frequencies. Across vowels, cisgender female formants were on average 17% higher than cisgender male formants. Some vowels were found to have larger scale factors than others, indicating larger differences between cisgender male and female
formants. The vowels selected as the stimuli for the present study, /ʌ/, /æ/, and /ɑ/, were chosen primarily for their relatively large scale factors, per Fant [14].

Previous studies have shown how distinctive acoustic patterns influence listeners’ perceptions of sexual orientation [15], as well as masculinity and femininity [16]. Such studies agree that the perception of gender in voice is influenced by both formant frequencies and F0, acting in conjunction with one another [11, 17, 18]. A study by Hillenbrand and Clark [8] revealed that manipulations that shifted both F0 and formant frequencies were usually effective in changing listeners’ perception of a speaker’s gender; however, shifting either one or the other alone was usually ineffective in altering perceived gender. With transgender women, increasing F0 alone has not been consistently effective in achieving increased feminine perception [5]. Furthermore, in a study that used visual analog scale (VAS) ratings to evaluate the relationship between F0 and transgender women’s happiness with their voice, increasing F0 alone was found to be unsuccessful in satisfying participants’ self-perception of femininity [19].

Few studies have explored the general efficacy of treatment or training methods aimed at increasing perception of vocal femininity in transgender women. In a longitudinal case study by Mount and Salmon [20], a 63-year old transgender woman received 11 months of voice modification treatment intended to increase F0 and modify F2 via altering tongue backness. The participant’s average F0 and F2 for vowels /i/, /a/, and /u/ significantly increased in post-treatment measures and were reported to result in an increased perception of femininity, although the methods of measuring this perceptual change were not delineated. Even fewer studies have explored the efficacy of intervention specifically targeting changes in formant frequencies. Carew, Dacakis, and
Oates [9] published a pilot study investigating the effectiveness of a treatment that aimed to raise F2 values with the goal of increasing perceived femininity in the voices of transgender women. In general, perceived femininity (as rated by outside listeners) tended to increase in post-session measures, as did the transgender participants’ self-perception ratings. Limitations of this study include a small sample size ($n = 5$), unspecified formant targets used during treatment, and results that were not sufficiently consistent across participants to be conclusive. More research on treatment efficacy for transgender voice modification is clearly needed and should consider further incorporation of formant manipulation.

Inspiration for formant manipulation in the present study comes from visual biofeedback. Visual biofeedback has been used to treat a variety of speech sound disorders, such as phoneme misarticulation [21, 22] and childhood apraxia of speech [23]. In the framework that analyzes speech interventions according to principles of motor learning [24], the success of biofeedback might be attributed to its real-time provision of “knowledge of performance” feedback, i.e. detailed qualitative information about a production attempt that can be compared to a model representing correct execution of the movement. In this study, participants received visual-acoustic biofeedback in the form of a dynamic display of formant frequencies via a real-time Linear Predictive Coding (LPC) spectrum (see Figure 1). Participants were asked to use the biofeedback to facilitate the modification of their speech to match several specified formant frequency targets. The formant target of primary experimental interest was a shifted-up version of each speaker’s average F2 frequency, intended to resemble the higher F2 frequency typical of a cisgender female speaker. As noted earlier, F2 was
chosen as the primary formant target because of its saliency as an acoustic indicator of gender.

Figure 1: A real-time LPC spectrum with superimposed line representing a target frequency for F2

The present research investigated several questions that need to be addressed before principles of biofeedback can be incorporated into treatment targeting increased femininity in the voices of transgender women. First, this study investigated the effectiveness of biofeedback training as a means to induce healthy adult speakers to deviate from their habitual formant targets. Given a shifted target frequency for a vowel, can speakers alter their production to match or approach the target? Second, this study explored the perceptual consequences of such an acoustic change. Would a formant shift in a given direction result in a change in blinded listeners’ ratings of the indexical properties of speech (e.g., an increased perception of femininity)? Or conversely, might the shifted vowels result in a different phonemic or phonetic interpretation of the acoustic signal (i.e., as a different or distorted vowel)? More specifically, would the trials targeting an increased F2 differ in femininity rating from neutral trials or trials targeting decreased F2?

2. METHODS

2.1
Participants

Two groups consisting of 12 transgender women and 19 cisgender men participated in the study. Transgender participants ranged from 21 to 71 years of age (M = 45.2; SD = 18.1), whereas the cisgender group’s ages ranged from 19 to 35 (M = 23.5; SD = 4.2). The use of a younger sample of cisgender males rather than an age-matched sample will be addressed, along with other limitations of the participant pool, in the Discussion section. Height, which influences vocal tract length and thus formant frequencies, ranged from 157.5 cm to 190.5 cm across groups. The transgender group’s mean height was 172.2 cm (SD = 10.7 cm), while the mean height of the cisgender group was 178.8 (SD = 8.4 cm). Speech and language history and demographic data for all participants are summarized in Table 1.

For inclusion in the cisgender group, individuals were required to be 18 years or older and be native speakers of American English with no history of major hearing, speech, language, or learning difficulties, per self-report. In addition, they were required to pass a pure-tone hearing screening at 500, 1000, 2000, 4000 Hz at 20dB HL, as well as a qualitative screening of speech, voice, resonance, and fluency based on a connected speech sample elicited with the Rainbow Passage [25]. Two cisgender participants were excluded: one on account of an atypical vocal quality and the other for being a British English speaker, both of which were revealed by the speech and voice screening. Two cisgender individuals who reported neurobehavioral conditions (e.g., ADHD, Tourette’s Syndrome) were included because their conditions were fully medically managed and were not judged to interfere with the completion of experimental tasks.
Inclusion in the transgender group was determined primarily on the basis of self-identification as a transgender woman (i.e., male-to-female transgender), per self-report. Each person was asked to select the group that they were electing to participate in (i.e. cisgender male, cisgender female, or transgender male-to-female), in addition to selecting their gender (i.e. female, male, transgender, other, or prefer not to answer). Due to the small size of the transgender population, it was necessary to include all candidates who reported this self-identification, even though this yielded a sample that was heterogeneous in several respects, including the wide age range noted previously. While not all participants elected to report their histories of voice modification, those that did reported varied experiences. Some participants had received previous voice modification therapy, some were either currently receiving hormone therapy or had previously; none had received laryngeal surgery. Participants varied in the amount of time they reported presenting as their “authentic selves” (i.e., their female gender presentation), ranging from never to 100% of the time. Not all participants elected to report this information. One transgender participant was a fluent speaker of English, but not a native speaker, and four individuals were included despite having failed the hearing screening at either 2000 or 4000 Hz, unilaterally or bilaterally. The potential effect that being a non-native English speaker may have had on phonemic production will be addressed in the Discussion.

This study was approved by the New York University Committee on Activities Involving Human Subjects (UCAIHS).

Table 1: Participants’ speech/language history and demographic data by group

<table>
<thead>
<tr>
<th>Spoken Languages</th>
<th>Transgender</th>
<th>Cisgender</th>
<th>Languages Spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speak only English</td>
<td>9</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Speak mostly English, but sometimes another language</td>
<td>Hebrew, Spanish, Turkish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speak mostly a language other than English</td>
<td>Spanish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speak English and another language in roughly equal amounts</td>
<td>Spanish, Portuguese, French, Cantonese, Oriya</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.2

Stimuli

The vowels /æ/, /ʌ/, and /ɑ/ were selected as the target vowels primarily because they represent front, neutral, and back vowels, respectively, and because they have relatively large scale factors as determined by Fant [14]. The vowels were put into the context of real words, surrounded by voiced plosives—*bud* (/bʌd/), *bad* (/bæd/), and *bod* (/bɔd/)—that were selected for their relatively neutral formant transitions.

Three target conditions were created for each vowel. The first condition featured a visual target for F2 that represented the speaker’s average baseline F2 for that vowel in the relevant word context, measured from three careful productions elicited during the initial baseline phase. This will be referred to as the “unshifted” condition. In the “shifted-up” condition, the speaker attempted to match a target with a scaled increase in F2, the derivation of which will be explained in the following subsection. This represents the condition of primary experimental interest, established to answer our question of whether increased F2 corresponds with an increase in perceived femininity. The third condition was the “shifted-down” condition, in which the participant tried to match a formant target lower than their natural F2 by the same amount. This condition was introduced to control
for the possibility that listeners might rate stimuli elicited under shifted conditions differently than the non-shifted stimuli for reasons unrelated to formant frequencies. For example, as an effortful speech task, matching a formant target could result in an elevated F0 [26] or could make speech sound unnatural.

The formula to compute formant targets in the two shifted conditions was individualized for each participant based on their mean F2 value. In an effort to maintain a constant level of task difficulty across participants, a fixed increment of shift was added to each participant’s mean F2. This fixed increment draws on Fant’s scale factors and a set of mean F2 values for Mid-Atlantic male speakers reported by Clopper et al. [27]. The size of the increment added differed across vowels due to the difference in the scale factors described by Fant [14]. The derivations of the target conditions for a given vowel are as follows:

1) Shifted-up target = F2_s + (F2_c x SF)
2) Shifted-down target = F2_s – (F2_c x SF)

where F2_s represents the speaker’s mean F2 (Hz) at baseline, F2_c represents the mean F2 from Clopper et al. [27], and SF represents the scale factor from Fant [14].

2.3

Procedure

All tasks were recorded in a soundproof booth with a Shure unidirectional microphone via a Marantz (24-bit encoding; 44,100 Hz sampling frequency) or Zoom (16-bit encoding; 16,000 Hz sampling frequency) recorder, including all target productions. This discrepancy in sampling rates is a consequence of experimenter error; however, it is considered to have little impact on the present measurements because no
phonemes distinguished by high-frequency elements (e.g., sibilants) were targeted. The Real-Time Linear Predictive Coding (LPC) function of the Sona-Match module of the Computerized Speech Lab (CSL; KayPENTAX) was used to provide the visual representation of each participant’s speech signal. The LPC filter order was customized for each participant, based on visual comparison of two candidate filter orders (23 and 25 peaks) at a sampling rate of 22 kHz.

Three tokens of *bud*, *bad*, and *bod* were recorded at the start of the session. To elicit these tokens, the experimenter verbally modeled clear productions of each word and displayed the orthography via a PowerPoint presentation. Transgender participants were instructed to use “whatever voice is most natural and comfortable” in an effort to elicit an accurate reflection of formant frequencies relative to the physical size of their vocal tract.

A trained student research assistant made immediate measurements of these baseline tokens in order to derive nine targets representing different vowel-condition combinations (e.g., *bud* in the shifted-up condition). For all three tokens of */bʌd/, */bæd/, and */bad/*, F2 was measured at the vowel midpoint using the acoustic analysis software Praat [28]. The LPC filter order was standardized at 5 formants in 5000 Hz across participants. The F2 values were inputted into a Microsoft Excel spreadsheet that calculated the mean across the three tokens of each word, as well as the shifted-up and shifted-down values using the standardized increment described previously. The order in which the nine vowel-condition combinations would be elicited for each participant was randomized using Excel.

Prior to the elicitation of the nine vowel targets, a brief introduction to biofeedback and context for the study was provided, as well as a generalized pre-practice
period allowing the participant to become familiar with the LPC biofeedback provided with the CSL Sona-Match program. The generalized pre-practice period did not have a set duration, lasting approximately 2-5 minutes depending on each participant’s own level of comfort. Participants were asked to focus only on manipulating the second peak, i.e. F2. They were instructed that in the subsequent portion of the study, their goal would be to manipulate the location of their “second peak” to match a specified target, indicated by the left edge of a ruler aligned with the appropriate target frequency on the LPC spectrum. Participants were informed of two strategies that they might use to manipulate the location of the peak: (a) changing the positioning of the tongue in the mouth (i.e., front or back), and (b) changing the shape of the lips (i.e., rounded or unrounded). Both groups were informed that altering pitch is not the preferred way to change the location of the peak. Transgender participants were asked to use their “best feminine voice” [5], if they had one, for all trials. In the generalized pre-practice period, participants were given a chance to get comfortable using the biofeedback display and to implement these strategies to shift the F2 peak to the left and right. The introductory script and instructions provided during this practice session are reported in Appendix A.

Following the generalized pre-practice session, each of the nine vowel targets was elicited in separate blocks that consisted of a block-specific pre-practice period, eight trials with biofeedback, and eight trials without biofeedback (i.e., the biofeedback program screen was minimized). All nine vowel-condition combinations are reported in Table 2; recall that the order of the blocks was randomized for each subject. The experimenter positioned a ruler along the LPC spectrum to align with a cursor generated by Sona-Match, indicating the F2 target frequency for a given block. Participants were
asked to match the highest point of the “second peak” with the left edge of the ruler as closely as possible. At the start of each block, participants were given up to three minutes to practice reaching the target frequency in a target-specific pre-practice period. Instructions for pre-practice were provided in a PowerPoint slideshow on a separate screen, and participants were asked to direct themselves through the hierarchy in a self-paced fashion (i.e., they could progress or return to previous levels as they saw fit). These instructions are also included in Appendix A.

Table 2: Nine vowel-condition combinations

<table>
<thead>
<tr>
<th>Block</th>
<th>Vowel</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ʌ</td>
<td>Shifted-up</td>
</tr>
<tr>
<td>2</td>
<td>ʌ</td>
<td>Shifted-down</td>
</tr>
<tr>
<td>3</td>
<td>ʌ</td>
<td>Unshifted</td>
</tr>
<tr>
<td>4</td>
<td>ɑ</td>
<td>Shifted-up</td>
</tr>
<tr>
<td>5</td>
<td>ɑ</td>
<td>Shifted-down</td>
</tr>
<tr>
<td>6</td>
<td>ɑ</td>
<td>Unshifted</td>
</tr>
<tr>
<td>7</td>
<td>æ</td>
<td>Shifted-up</td>
</tr>
<tr>
<td>8</td>
<td>æ</td>
<td>Shifted-down</td>
</tr>
<tr>
<td>9</td>
<td>æ</td>
<td>Unshifted</td>
</tr>
</tbody>
</table>

Once either the participants were satisfied with their manipulation of the second formant to reach the target or three minutes had expired, they were instructed to produce the target word (i.e. *bud, bad, or bod*) for eight consecutive trials while matching the target as closely as possible. Immediately after, the biofeedback was removed and participants were asked to say the word eight more times the exact same way as they had said it with the biofeedback. These no-biofeedback trials were elicited because utterances produced while using biofeedback have the potential to be artificially elongated, which could alter listeners’ perceptual ratings.

2.4
Measurements

Praat acoustic analysis software [28] was used to obtain acoustic measures of all vowel-condition combinations in both biofeedback and no-biofeedback contexts. The acoustic measurements obtained were the first three formant frequencies, F1, F2, and F3, as well as the fundamental frequency, F0. Trained student research assistants measured the participants’ formants by listening to the audio recording of each word while viewing a spectrogram with automated LPC tracking of the first five formants. The standard LPC filter settings for cisgender male vocal tracts (five formants in 5000 Hz) were used across participants, with the exception of two transgender participants, whose acoustic profiles matched more closely with the default formant settings for cisgender female vocal tracts (five formants in 5500 Hz). The student research assistants identified the vocalic portion of each word-level utterance and placed the cursor at the midpoint of F2. Points that appeared to be outliers relative to adjacent points in the automated formant track were avoided. Lastly, an automated algorithm was used to extract the measurements of F1, F2, F3, and F0 from the selected point in each word.

Reliability was calculated by collecting a second student’s measurements for approximately 20% of all measurements. The two sets of measurements of F1, F2, and F3 values were calculated to be correlated with Pearson’s $r = 0.98$, indicating a high level of agreement. All items for which the difference between original and reliability measures fell more than three standard deviations away from the mean difference across the full set of re-measured tokens were manually examined ($n = 11$). Cases where the original measurement appeared to be more accurate were retained, and cases where the original measurement appeared to be in error were discarded ($n = 1$).
Once all formant measurements were collected, the subjects were divided by group (i.e., cisgender and transgender) and the dataset was cleaned for outliers. Outliers were identified as tokens that fell at least 2.5 standard deviations above or below the mean midpoint values of F0 and F2 for a given subject. A total of 76 outliers were identified within individual subject data and were excluded from further analyses. We also investigated whether any individuals fell at least 2.5 SD away from the mean F0 or F2 for their group (i.e., cisgender or transgender), but no subjects were identified as outliers with respect to this criterion. Token-to-token variability was also calculated and compared across subjects within each group. One subject in the cisgender group was found to be an outlier with respect to F0 variability. However, the cisgender group’s average token-to-token variability was quite low (see Results), and this individual’s data were retained as falling broadly within the range of normal variation.

2.5

Perceptual Rating

Perceptual ratings of the femininity of speech tokens elicited in the study were obtained from blinded naïve listeners. Only the tokens produced in the no-biofeedback context were presented for rating in order to avoid altering listeners’ perceptions with potentially elongated productions, as mentioned previously. Perceptual raters were recruited through the online crowdsourcing platform Amazon Mechanical Turk (AMT). Stimuli for perceptual rating were presented and responses were recorded using the Experigen online platform [29]. AMT has been utilized to facilitate the collection of data from large numbers of experimental subjects in the fields of behavioral psychology [30, 31], linguistics [32, 33], and, recently, speech and language pathology [34, 35].
McAllister Byun et al. [36] explored the validity of AMT listeners’ ratings of child speech by comparing them to the ratings of experienced listeners. Strong agreement between experts and crowdsourced listeners was found when responses were aggregated across large groups of listeners, and agreement remained adequate when using a sample size small enough to be considered practical for use in experimental measurement (i.e., $n = 9$ listeners).

A follow-up study by McAllister Byun, Harel, Halpin, & Szederi [37] found a high level of agreement between binary ratings and ratings on a VAS when results were aggregated across multiple crowdsourced listeners. In visual analog scaling, raters are presented with a linear scale with defined endpoints and are instructed to select any point along the line to represent the prototypicality of each token relative to those endpoints. Following Munson et al. [16], listeners in the present study were instructed to rate speakers’ gender typicality on a VAS with “definitely male” at one endpoint and “definitely female” on the other (Figure 2). Raters were asked to select a point on the scale to convey their degree of confidence in their ability to identify the speaker’s gender based on the single word presented. The specific instructions provided to the raters for using the VAS were adapted from Julien & Munson [38] and are reported in Appendix B. Each token presented was accompanied by an orthographic representation of the word, and raters were able to listen to each token up to three times. Tokens were presented in blocks of approximately 200, and raters were allowed to complete multiple blocks.

Figure 2: Visual Analog Scale (VAS) used for stimulus rating.
AMT raters were naïve to the purpose of the study and to the inclusion of transgender speakers in the study sample. The productions that were rated were elicited from both cisgender male and transgender women groups in the no-biofeedback context. In order for listeners to hear a full range of variation from highly masculine to highly feminine voices, “filler” tokens produced by cisgender females were additionally presented. Cisgender female productions of each word (e.g. bud, bad, and bod) were elicited from 10 student assistants. Each cisgender female speaker produced 30 tokens of each word, resulting in a total of 600 different tokens. These “filler” tokens were repeated to comprise one third of all presented productions, with the experimental productions comprising two thirds, in order to adequately balance transgender women, cisgender male, and cisgender female speakers.

Responses were collected until each token had been rated by 9 unique listeners. We chose to aggregate over 9 listeners based on McAllister Byun et al. [36]’s finding that this number of naïve listeners yielded responses comparable to responses aggregated over 3 trained listeners. However, those findings pertained to AMT listeners’ ratings of the accuracy of child speech and cannot necessarily be assumed to transfer to ratings of the gender typicality of adult speakers; this is a limitation of the present study. To increase the reliability of these ratings, prior to completing experimental rating blocks, listeners
were required to pass an eligibility-testing block measuring the reliability with which they used the VAS to rate speech tokens. Following a protocol developed in previous research on the use of repeated ratings to identify high-performing individuals among crowdsourced listeners [39], a sample of 30 tokens were repeated four times in random order, totaling 120 tokens. Raters were excluded if their intraclass correlation coefficient (ICC) across the repeated ratings was lower than 0.8. Raters were instructed to wear headphones and were required to report the brand of the headphones used. A total of 26 unique raters passed the eligibility test and contributed to the ratings reported below.

3. RESULTS

3.1 Acoustic Measures

Because perceptual ratings were only collected for the no-biofeedback trials, only those data will be included in the analyses that follow. A t-test revealed that these two contexts did not differ significantly with respect to F0 (biofeedback mean 126.14, SD 31.92; no-biofeedback mean 124.68, SD 30.12; \( t = 1.56, df = 4371.49, p = 0.12 \)). They also did not differ with respect to F2 (biofeedback mean 1370.89, SD 270.04; no-biofeedback mean 1371.87, SD 272.78; \( t = -0.12, df = 4383.06, p = 0.91 \)). Therefore, these results can be regarded as reasonably representative of both biofeedback and no-biofeedback trials.

In the following boxplots, the middle bar represents the median value, the box represents the interquartile range (25th to 75th percentile), and outliers are represented as single points. Figure 3 shows the distribution of F0 values measured in each vowel-
condition combination for both cisgender male and transgender female groups, while
Figures 4 and 5 show the corresponding distributions of F1 and F2 values, respectively.

Figure 3 shows that, overall, the transgender female group produced higher F0 values than the cisgender group. In addition, the transgender female group exhibited much higher F0 variability than the cisgender male group across all vowels and conditions. Within each group, visual inspection of the boxplots does not indicate any reliable pattern of F0 differences across vowels or conditions.

Figure 3: Boxplots representing F0 frequencies by vowel (represented by orthographic target bad, bod, or bud), condition, and group (cisgender male, transgender female).

Figure 4 shows that F1 values produced by the transgender female group tended to be more variable than those produced by the cisgender male group. The three vowels elicited are only partly distinguished by F1, with the /ʌ/ in bud associated with a lower F1 than the /æ/ in bad and the /ɑ/ in bod. Some differences were observed between shifted-
up, shifted-down, and unshifted target conditions, but these differences were generally small in magnitude and showed no clear pattern across vowel targets and groups.

Figure 4: F1 frequencies by vowel (represented by orthographic target *bad, bod, or bud*), condition, and group (cisgender male, transgender female).

In contrast with the preceding figures, Figure 5 shows that the target formant, F2, differed consistently across elicitation conditions. F2 values produced in the shifted-up condition were highest, while F2 values produced in the shifted-down condition were lowest, relative to the unshifted condition. This pattern was present across all vowel targets and across groups.

Figure 5: F2 frequencies by vowel (represented by orthographic target *bad, bod, or bud*), condition, and group (cisgender male, transgender female).
A linear mixed-effects regression was fitted over productions elicited without biofeedback from all subjects, vowels, and conditions. As F2 is the primary variable of experimental interest, the F2 frequency in Hertz, measured at vowel midpoint, served as the dependent variable. By-subject random intercepts and slopes were examined because data points were nested within subjects. The fixed effects of primary interest were condition (shifted-up, unshifted, shifted-down), group (cisgender male, transgender female), and vowel (/ʌ/ in *bud*, /æ/ in *bad*, and /ɑ/ in *bod*). A fixed effect of F0 at midpoint was included to control for any influence of vocal pitch. Model selection was performed using likelihood ratio tests, and only those interactions and random slopes that yielded a significant difference in likelihood relative to a minimally reduced model were retained. The final model emerging from this process included the interaction between
condition and vowel and by-subject random intercepts with crossed random slopes for condition and word.

In the selected model, there was a significant effect of group ($\beta = 94.14, SE = 28.87, p = 0.0026$), with the positive coefficient indicating that F2 values produced by the transgender female group were generally higher than those produced by the cisgender male group. A significant effect of condition was additionally observed: the shifted-down condition was associated with significantly lower F2 values than the unshifted condition ($\beta = -104.74, SE = 30.68, p = 0.0019$), and the shifted-up condition was associated with significantly higher F2 values than the unshifted condition ($\beta = 57.98, SE = 19.46, p = 0.0057$). There was a significant effect of vowel, whereby the value of F2 in /æ/ differed significantly from both /ɑ/ ($\beta = -521.17, SE = 33.05, p < 0.001$) and /ʌ/ ($\beta = -400.26, SE = 32.76, p < 0.001$). Vowels /ɑ/ and /ʌ/ also differed significantly in their F2 values ($\beta = 120.9, SE = 21.93, p < 0.001$). Finally, a significant interaction between condition and vowel was retained in the final model because it made a significant contribution in a likelihood ratio test ($\chi^2 = 10.89, p = 0.028$). However, this interaction did not lend itself to an obvious interpretation. Full fixed-effects from this and all other models are provided in the online supplemental materials.

The same model was fit with formants that were not explicitly targeted, F1 and F3, as the dependent variables. In the model that examined F1, there was still a significant effect of condition such that the shifted-down condition featured significantly lower F1 values than the unshifted condition ($\beta = -30.61, SE = 11.17, p = 0.01$), although the unshifted condition did not differ from the shifted-up condition ($\beta = 21.50, SE = 12.72, p = 0.10$). There was no significant effect of group in the model examining F1. In
the model that examined F3 as the dependent variable, there was no effect of condition; however, there was a significant effect of group ($\beta = 181.6, SE = 52.95, p = 0.0016$), with the positive coefficient indicating that F3 values produced by the transgender group were generally higher than those produced by the cisgender male group. Lastly, a model with F0 as the dependent variable showed no effect of condition, but the effect of group was again significant, with the transgender group producing higher F0 values on average than the cisgender male group ($\beta = 31.03, SE = 6.65, p < 0.001$). See online supplemental materials for complete results of these comparison models.

3.2 Perceptual Ratings

A variable representing average perceptual rating, aggregated across nine unique listeners per token, will be referred to as “mean click location.” Mean click location values range from 0 to 1, with 0 representing the “definitely male” end of the VAS and 1 representing the “definitely female” end. Thus, higher values correspond with greater perceived femininity.

Overall, the transgender female group received higher mean click locations than the cisgender male group, as seen in Figure 6. There was also a higher degree of variability in the femininity ratings assigned to transgender female speakers. The median of the distribution of mean click locations for the transgender speaker tokens, pooled across words and conditions, was 0.17. While this is notably higher than the median of 0.05 that was calculated for the cisgender male group, it still fell well below the midpoint of the scale. Thus, transgender female speakers were still generally perceived as male.
For comparison, the filler tokens elicited from cisgender female speakers received a median click location of 0.95. (These filler tokens are not depicted in Figure 6 because they were not elicited in the context of a biofeedback-matching task and thus cannot be subdivided by condition.)

A subset of tokens from the transgender female group did fall in the upper quartile of the scale, indicating that the average listener perceived these tokens to have been produced by female speakers. It was determined that two speakers in the group were responsible for the great majority of these tokens. For this reason, the analyses reported below were conducted both with and without these speakers.

Figure 6: Mean click location by vowel (represented by orthographic target bad, bod, or bud), condition, and group (cisgender male, transgender female).

Ratings of perceived femininity are plotted in relation to acoustic measures in Figure 7. F2 is represented on the x-axis, mean click location is represented on the y-axis,
and $F_0$ is represented in the graded shading of each point, with lighter colors corresponding with lower $F_0$ values and darker colors with higher $F_0$ values. (This inverse relationship between $F_0$ and darkness is adopted for visibility, since low $F_0$ values are more plentiful and more densely packed than high $F_0$ values.) As seen in the preceding analysis of acoustic data, the data from transgender female group extends into higher values in both the $F_0$ and $F_2$ dimensions.

Figure 7: Perceptual rating of tokens by group (cisgender male, transgender female) and $F_2$ frequency. $F_0$ is represented by color, with darker colors representing higher $F_0$ values.

The significance of the relationship between acoustic measures and ratings of perceived femininity was examined in a mixed-effects linear model with mean click location as the dependent variable. Independent variables were the acoustic measures of $F_0$ and $F_2$, as well as group (cisgender male versus transgender female) and word (bad,
bod, bud). As previously, only those interactions and random intercepts and slopes that made a significant contribution to model likelihood were included in the final model. Following this criterion, the final model included an interaction between the acoustic variables of F2 and F0, as well as a random effect of subject with random slopes for word, F0, and F2.

There was a significant effect of group ($\beta = 0.16, SE = 0.06, p = 0.0083$), indicating that tokens produced by the transgender group were generally rated more feminine than those produced by the cisgender male group. There was also a significant effect of F2 ($\beta = 0.02, SE = 0, p = 0.002$), with the positive coefficient indicating that tokens with higher F2 values tended to be rated more feminine than tokens with lower F2 values. The effect of F0 was likewise significant with a positive coefficient, indicating that tokens with higher F0 values were rated more feminine than tokens with lower F0 values ($\beta = 0.05, SE = 0.01, p < .001$). A significant effect of word indicated a difference in femininity rating between bad and bod ($\beta = 0.03, SE = 0.01, p = 0.0054$), but the underlying basis driving this difference was not immediately clear. Finally, a significant interaction between F0 and F2 suggested that these two acoustic variables made a joint contribution, rather than an independent contribution, to listeners' femininity ratings ($\beta = 0.01, SE = .002, p < .001$). See online supplemental materials for complete results of these models.

It was previously noted that two individuals in the transgender group had received the overwhelming majority of ratings in the highest (most feminine) quartile of the VAS. To determine if these subjects were contributing disproportionately to the observed effects, the model was re-fitted with these two individuals excluded. All of the same
effects and interactions remained significant in the model fitted to the adjusted data set. Complete results of this model are reported in the online supplementary materials.

As described previously for the models analyzing acoustic changes, we examined how model results were affected if the targeted formant, F2, was replaced with formants that were not explicitly targeted, F1 and F3. The structure of the model was held constant apart from the substitution of F1 or F3 for F2 in both fixed and random effects terms. The model examining F1 yielded no significant effect of F1 on femininity rating ($\beta = 0.007$, $SE = 0.005$, $p = 0.17$), although there was a significant interaction between F1 and F0 ($\beta = 0.009$, $SE = 0.0027$, $p = 0.001$). The model examining F3 likewise showed no significant association between F3 and femininity rating ($\beta = 0.004$, $SE = 0.003$, $p = 0.23$) but did yield a significant interaction between F3 and F0 ($\beta = 0.0089$, $SE = 0.0024$, $p < 0.0001$). Complete model results can be found in the online supplementary materials.

4. DISCUSSION

The present study investigated the feasibility of using biofeedback in voice modification therapy for transgender women by asking two questions: (1) Can speakers use biofeedback to manipulate their F2 frequency during word production to match a target formant frequency typical of a cisgender female? (2) If so, does this acoustic shift correspond with a change in the speaker’s perceived femininity?

4.1 Acoustic Changes

Both transgender female and cisgender male groups successfully altered their F2 values using visual biofeedback. A linear mixed-effects model revealed that participants
produced statistically significant differences in F2 values when viewing visual targets that were shifted up, shifted down, or unshifted relative to their baseline F2 for a given vowel. In other words, despite the limited nature of the training and practice they received, participants were generally successful in shifting their F2 frequencies in the direction of a target with visual feedback.

Besides the changes in F2 that were targeted through biofeedback, additional mixed-effects models revealed condition-dependent changes in parameters that were not explicitly targeted. Although only F2 was explicitly targeted through biofeedback, F1 frequency showed an effect of condition that mirrored that of F2: relative to the unshifted condition, F1 was significantly higher in the shifted-up condition and significantly lower in the shifted-down condition. This resembles a result reported in the study by Carew, Dacakis, and Oates [9], where transgender women speakers who were trained to increase F2 also showed a significant increase in F1, even though that formant was not targeted. In the context of biofeedback, a novel task, participants are unlikely to be able to control the frequency of a single formant with a high level of precision. It is thus not surprising that efforts to raise F2 sometimes yielded a broader shift in the overall concentration of energy across frequencies, raising F1 as well. It is also possible that some speakers achieved an increase in F2 frequency partly by raising the height of the larynx. This would have the effect of shortening the length of the vocal tract resonator and thus could be expected to raise all formants. However, no effect of condition was observed for the third formant F3, nor for the fundamental frequency F0.

With respect to between-group differences, F0 was unsurprisingly found to be higher in the transgender female group, who were instructed to speak in whatever they
considered their “best feminine voice.” In addition, the transgender group was observed to produce higher average values for F2 and F3. Both F2 and F3 are increased by lip rounding and decreased by lip spreading. It is possible that transgender speakers tended to use a more open lip posture in an effort to adopt a more feminine speech style. This strategy could have been explicitly trained in some of the participants who had received or were receiving voice modification therapy.

Across all vowels and conditions, the transgender female group exhibited higher variability of F0, F1, and F2 than the cisgender male group. These group-level differences in variability are likely to be attributable to the high level of between-participant heterogeneity present in the transgender female group, where participants had differing exposure to voice modification therapy and may also have differed in their interpretation of the instruction to use their “best feminine voice.” Participants in both groups showed a wide range of individual variability in how successful they were in shifting their formants to match the biofeedback targets. It is possible that these differences in target-matching could be attributed to individual differences in other domains, such as degree of motivation, capacity for sustained attention, or motor skill in adjusting the configuration of the vocal tract. Finally, some speakers may not have adequately understood the nature of the task on the basis of the standard level of biofeedback training provided, which was relatively brief. It is possible that some participants would have benefitted from a more extensive training period.

4.2

Perceptual Rating
Previous literature has shown that F0 is the most salient acoustic indicator of gender in speech [2, 4-6], but formants also play a role in gender perception [5, 9, 11]. Ultimately, there is consensus that F0 and formant frequencies—and F2 in particular [13]—act in conjunction with one another to influence the perceived gender of a speaker [5, 8].

The present study supports previous research in finding a significant influence of F2 frequency on perceived gender. The mixed-effects model showed a significant association such that tokens with higher F2 values received higher perceptual ratings of femininity. In other words, voices with higher F2 values tended to be perceived as more feminine. The model also yielded a significant effect of F0, indicating that voices with higher F0s were associated with higher femininity ratings. Crucially, the best-fitting model included not only main effects of F2 and F0, but also the interaction between them. This agrees with previous studies reporting that both F0 and F2 contribute significantly to gender perception, and that there is a complex relationship between these two parameters and perceived gender. When comparison models were fitted with F1 or F3 instead of F2 as the formant frequency used to predict femininity rating, the interaction with F0 remained significant, but the non-targeted formants did not make a significant contribution on their own. This supports previous research reporting that F2 plays a particularly important role in influencing listeners’ judgments of speaker gender [13].

In general, transgender female speakers were perceived as sounding more feminine than cisgender male speakers. This is not surprising given that transgender speakers were encouraged to use their “best feminine voice,” were motivated to sound
feminine, and in some cases had previously received training in altering their pitch and/or formants to achieve a more feminine percept. However, while the transgender female group was found to have obtained a higher median click location than the group of cisgender male speakers (i.e., closer to the “definitely feminine” end of the VAS), the transgender group’s median rating still fell below the midpoint of the scale, closer to the “definitely male” label than the “definitely female” label. Thus, the transgender female voices still ultimately tended to be perceived as male by blinded listeners.

4.3

Limitations and Future Directions

Several limitations to the study should be noted. Our recruitment efforts yielded only 12 participants in the transgender female group. Due to the small size of the overall population of transgender women, all candidates who opted to participate in the study were included. This resulted in a highly heterogeneous sample, including a wide age range and a varied history of experience with voice modification therapy. Additionally, inclusion in the transgender group allowed for one speaker who, while fluent in English, was a native Spanish speaker. It is possible that the acoustic properties of this speaker’s vowels were influenced by a L1 Spanish L2 English dialect. Given the heterogeneity and the small number of participants, it is unclear whether the group-level results reported here can be expected to generalize to the broader population of transgender women.

Productions were elicited at the single-word level in order to simplify the task of target matching via biofeedback. This study has not tested whether biofeedback can be used to shift acoustics in a more naturalistic context (i.e. connected speech vs. isolated
single words), nor whether such shifts would be associated with changes in perceived femininity. Furthermore, this study has not assessed whether training in biofeedback would carry over outside of the clinical treatment setting. Future studies investigating the effectiveness of target matching within the context of connected speech would allow for more naturalistic assessment of the impact of biofeedback training on perceived femininity. However, it is highly probable that a much more prolonged period of training would be necessary before generalization to connected speech can be expected. Furthermore, it is important to note that this study did not account for the role of speaker gender cues other than F0 and formants in gender perception, such as intonation, vocal quality, articulation, word choice, and non-verbal communication (e.g., gestures). Because these elements have been found to influence perceived gender, a comprehensive program of voice modification therapy would address all of these aspects.

Another limitation of this study arises from the fact that raters knew the target word that was being presented in each case, since audio tokens were accompanied by orthography. Without this context for each production, it is possible that raters could have perceived the words produced with shifted formants as different words than what was elicited. For instance, a production of *bad* with a high F2 frequency might be interpreted as a feminine speaker’s production of *bad* in the context of orthographic support, but in the absence of orthography, it might be interpreted as a more masculine speaker’s production of *bed*. The vowel change associated with a shift in F2 would result in a real word in some cases and a non-word in others, further complicating the interpretation of tokens presented without orthography. A useful follow-up study would be to investigate whether the gender ratings assigned to the current speech samples would differ
significantly if items were presented without an orthographic representation of the target word.

The vowels used as targets were selected primarily because of their large scale factors, i.e. sizable differences in typical formant heights between cisgender male and female speakers [14]. For vowels with less salient scale factors, biofeedback to shift formants might be expected to have limited effect on perceived gender. Additionally, vowels with intrinsically high F2 values (e.g. /i/) were not targeted because it might have been more difficult and perhaps unattainable for participants to match a markedly raised F2 target when they are already at the upper bound of their F2 range.

Finally, it should be emphasized that the influence of F2 on perceived femininity cannot truly be separated from the influence of F0 in our sample, even though both F0 and F2 were seen to make a significant contribution as main effects in the best-fitting mixed-effects model. Two transgender female speakers who had both a high F0 and a high F2 were rated as more feminine than other participants. (Recall, though, that the same main effects and interactions remained significant when the model of perceived femininity was re-fitted with these two individuals excluded.) Because these individuals were successful in adjusting both F0 and F2 in their acoustic output, it is not clear which factor had a larger influence on listeners’ ratings of perceived femininity. The role of F2 would be more readily isolated with a larger, more homogeneous sample of participants. As an alternative, future studies could better control for F0 by systematically manipulating F0 in tokens presented to raters. F0 could either be manipulated during production using an electrolarynx, or it could be synthetically altered post-recording. In
either case, exercising experimental control over F0 would allow for a better understanding of the specific contribution of F2 to perceived femininity.

Overall, this study provided a proof-of-concept demonstration that visual-acoustic biofeedback might be a feasible means to manipulate formant heights and thereby influence perceived gender. Future studies should move towards research on the efficacy of biofeedback training in the context of transgender voice modification therapy. For instance, the magnitude of changes in perceived femininity could be compared across biofeedback and non-biofeedback conditions in a single-subject experimental study in which biofeedback training would be provided over a more extended period of time.

5. CONCLUSION

This study investigated the feasibility of using biofeedback to increase perceived femininity in the speech of transgender women. In general, participants were successful in manipulating the height of F2 using visual-acoustic biofeedback, and higher F2 values were associated with higher ratings of the perceived femininity of speech. Consequently, this study offers a preliminary suggestion that biofeedback could be a useful tool in voice modification therapy for transgender speakers. However, studies with larger sample sizes and more information about generalization will be necessary in order to evaluate the clinical utility of the technique. The present study confirms previous findings that F0 and F2 jointly contribute to influence listeners’ perception of speaker gender.
APPENDICES

A. Introductory Script and Generalized Pre-Practice Instructions & Target-specific Pre-practice Slideshow Instructions

Introductory Script and Generalized Pre-Practice Instructions

Introduction
- “The peaks represent ‘vowel frequencies’ and correspond with each individual’s vocal tract shape, which is influenced by things like your tongue position and the shape of your lips.”
- “Where the peaks are located depends on the vowel sound being produced, which is created by changing the positioning of your tongue and lips. So, the locations of the peaks will be different for each vowel.”
  - Model point vowels (/u/, /i/, /ɑ/, and /æ/)
- “Not only do the peaks differ for each vowel, as you can see, but there is variability within a given speaker for a single vowel, depending on how the lips and tongue are positioned. This will correspond with differences in peak locations.”
  - Demonstrate /ʌ/ with rounded and unrounded lips
    - “For example, you could say ‘uh’ with rounded lips or unrounded lips.”
    - “See how the peaks change depending on what I’m doing with my lips?”

Focusing on F2
- “For this study, we’re going to be focusing on the second peak” [point to second peak]
- “Focus on where the highest point of the peak is located on the x-axis, which shows frequency” [point to x-axis]
- “Some things to note: The number of peaks can fluctuate, and sometimes F1 and F2 merge”

Pitch is not a factor in biofeedback
- “The placement of the peaks do not depend on pitch. So, the peaks stay steady while pitch is shifted, and the peaks change position will the pitch stays the same.”
  - Demonstrate: increase and decrease pitch of /u/
  - Demonstrate: move lips and tongue, keeping pitch the same
- “So, changing the position of your tongue and lips is more important than changing the pitch of your voice to manipulate the peaks.”
• For transgender participants only: “To keep pitch consistent throughout, please use your preferred feminine pitch throughout all tasks. Since we are not measuring pitch, the pitch you use will not affect the results.”

Context for the study
• “I will be presenting 3 vowels with 3 different target frequencies. I will ask you to produce each word and try to match the frequency we’re presenting. I want to know: could matching a target frequency make a vowel sound more feminine?”
• Explain how each vowel will have a target, indicated by the ruler. The objective is to make the highest point of the second peak only line up with the line. Repeat and explain as necessary.
  o “Each vowel will have a target, which will be indicated by the placement of the ruler. The goal is to make the highest point of the second peak only line up with the left edge of the ruler. Does that make sense?”
• Explain that some of the targets will be easy, and some of them will be more difficult.
  o “Some of the targets may be easy for you to hit, and some may be more challenging. I want you to be persistent—try a bunch of different things to match the target before giving up.”

Generalized pre-practice session
• “Now we’re going to look at 3 different vowels using the biofeedback, and you can have a chance to experiment with it.”
• “Two tricks for moving the second peak around are by rounding or unrounding your lips, as we looked at before, and moving your tongue forward and back.”
• Practice the 3 target vowels with biofeedback (in isolation and in the context of a word) and then without biofeedback:
  o “Now we’re going to practice this with the 3 vowels that we’ll be looking at in this study. We’ll start with the vowel /ɑ/ in isolation. Focus on moving the second peak up and down.”
    ▪ Give general cue to move the second peak up and down, since specific target location will not be available yet.
  o “Now try saying the vowel in a word, while watching the biofeedback. See how you can move the second peak up and down. Say ‘bod’.
  o “Now try saying the word the same way as you just did, with your eyes closed and without looking at the biofeedback. Just go by feel.”
• Repeat all steps for /ʌ/ and /æ/

Target-specific Pre-practice Slideshow Instructions

Slide 1
Practice just the vowel sound.
When you have matched the target as closely as you can, click the right arrow to move on.
Slide 2
Practice the word, stretching out the vowel.
When you have matched the target as closely as you can, click the right arrow to move on.

Slide 3
Practice the word, making it sound more natural.
Can you close your eyes and hit the target, just by feel?
When you have matched the target as closely as you can, click the right arrow to move on.

Slide 4
Are you sure you have matched the target as closely as you can?
Click the left arrow to go back for more practice at any level.
B. Perceptual Rating Instructions

This study investigates adults’ perceptions of gender within adult speakers. You will see three written words (“bud”, “bad”, and “bod”), presented one at a time, and you will hear an adult saying that word. Please listen carefully to determine whether you think the speaker is male or female. You can click to listen to each file up to three times.

Your task is to rate the person’s gender based on their production of one of the three words. This rating will be made on a line. One end of the line says “definitely male.” The other end says “definitely female.” When you hear what you think is a DEFINITELY MALE speaker, click on the line where it says “definitely male.” When you hear what you think is a DEFINITELY FEMALE speaker, click on the line where it says “definitely female.” Sometimes, you won’t be sure whether the speaker is definitely male or female. In those cases, you should click somewhere in between the two ends of the line. If you thought the speaker sounded more male than female, click somewhere on the line closer to the text that says “definitely male.” If the speaker sounded more female than male, click somewhere on the line closer to the text that says “definitely female.”

We hope that you will use the whole line when rating these sounds. We don’t have any specific instructions for what to listen for when making these ratings. We want you to go with your gut feeling about what you hear.
ACKNOWLEDGEMENTS

The authors would like to thank all of the participants in the study for their generous contribution. We would also like to thank Daphna Harel for statistical advice, Daniel Szeredi for coding assistance, Darlene Monda for help with recruitment, and all of the speech-language pathology students from the Biofeedback Intervention Technology for Speech Lab (BITS Lab) at NYU Steinhardt for their help with recruitment, data collection, and acoustic measurement.
REFERENCES


VITAE

**Tara McAllister** received A.B. and A.M. degrees in linguistics from Harvard University, Cambridge, MA (2003), an M.S. in speech-language pathology from Boston University, Boston, MA (2007), and a Ph.D. in linguistics from the Massachusetts Institute of Technology, Cambridge, MA (2009). Since 2012, she has worked as an Assistant Professor in the Department of Communicative Sciences and Disorders at New York University. Dr. McAllister’s research on technology-enhanced intervention for treatment-resistant speech errors is supported by the National Institute on Deafness and Other Communication Disorders (National Institutes of Health).

**Deanna Kawitzky** received a M.S. degree in communication sciences and disorders from New York University, New York, NY (2016). She became a member of Dr. McAllister’s Biofeedback Intervention Technology for Speech Lab (BITS Lab) at NYU Steinhardt in 2013, where she conducted the present research as her Master’s Thesis. This research was delivered as a poster presentation at the 2016 American Speech-Language-Hearing Association (ASHA) conference. She currently works at Burke Rehabilitation Hospital in White Plains, NY.

REFEREES

- Adrienne B. Hancock, George Washington University, hancock@gwu.edu
- Marylou P. Gelfer, University of Wisconsin—Milwaukee, gelfer@uwm.edu
- Sally Gallena, Loyola University, sgallen@loyola.edu
- Mary Gorham-Rowan, Valdosta State University, mmgorhamrowan@valdosta.edu