The Male-to-Female Transgender Voice: Most Salient Voice Parameters in Perceived Gender Identification

Emily Manny
Northeastern University—Boston, Massachusetts, USA

Summary:
This review considers the question: what are the most salient vocal parameters to perception of a male-to-female transgender speaker as a passing female? Much of the current research suggests that speaking fundamental frequency must reach a certain level for the speaker to be perceived as a female by a naïve listener. If the speaker’s fundamental frequency lies in a determined ambiguous range, other parameters become more important cues to speaker gender. The next most-studied parameter is that of the first three formants of a speaker’s vowels, F1, F2, and F3. Studies suggest that biologically female resonance characteristics that can be achieved through therapy allow a speaker to produce the higher vowel formants associated with female speech. Nearly all research points to the possible implications of breathiness and intonation to gender perception as well, and while these are not as thoroughly researched, current findings imply that they also make minor contributions to a female-sounding voice. When standard measures for breathiness and intonation can be identified, researchers will more easily be able to support this claim with the importance of a certain level of breathiness or intonation to the perception of a speaker as a female.

Introduction:
So many biological males and females take the gendered parameters of their voices for granted. Though each person’s voice “sounds” different, for a person to produce a voice that sounds inherently “male” or “female” is often as instinctive as breathing, requiring no thought at all. For transgender people, gendered voice production can be a long process and an ongoing challenge. An authentic “male” or “female” voice is essential to projecting the image of gender he or she feels on the inside to the outside perception of this person as male or female by others. Fear of misidentification and the social isolation this problem may lead to is common among transgender folk; hence, they often seek assistance with their voice change process. Female-to-male patients tend to have an easier voice transition due to permanent changes in the vocal fold anatomy as a result of testosterone therapy\(^1\). Because male-to-female, or MtF, patients most often do not achieve this adequate voice change through hormone therapy\(^2\), they typically seek the help of speech language-pathologists who can assist them with the gendered voice transition through voice therapy, and this review will focus on the voice changes of this population. Ultimately, the goal for these clients is to be perceived as female by listeners. In order to ensure that clients achieve such goals, the speech pathologist administering therapy must address those voice parameters most salient in determining a listener’s perception of the voice as biologically male or female.

Nearly all related studies agree that overall pitch, or speaking fundamental frequency (SFF, \(f_0\)), is the most crucial contributing factor to gender identification of the speaker. Still, it is also important to note that increasing SFF to the crucial level
is most often not enough for the client to be consistently perceived as a female. Wolfe et al. is cited by Hancock et al. and many other current investigators as having conducted a gender perception study in which they discovered that one of the MtF voices that was judged by listeners to be the most “feminine” voice in the group had the lowest fundamental frequency. On a similar note, Owen and Hancock report that Mount and Salmon first discovered in 1988 that although their MtF client had reached the crucial SFF at 210 Hz, she was not perceived on the phone as a female until her vowel formants increased.

Current research agrees that although fundamental frequency clearly plays a crucial role, it cannot stand alone as the sole determining voice parameter of gender perception. Gelfer and Bennett also revisit the work of Wolfe et al. in their reference to a boundary frequency region between 145 Hz and 165 Hz in which some voices are perceived as male and others are perceived as female. All voices below are perceived as male and all above as female, but for voices in this middle zone, where many transgender speakers find themselves, the aforementioned increased vowel formant frequencies could become important cues to the perception of gender, depending on the type of utterance (sentence, word, syllable, etc.) being studied. Still, there are additional components to the voice that serve as perceptual cues for listeners. The findings of Hillenbrand and Clark acknowledge that SFF and vowel formants alone do not provide enough cues even together to consistently distinguish the difference between a male and female voice. They invoke the consideration of additional voice cues such as prosodic features of the voice, as well as breathiness. While these are lesser researched areas, there is evidence that the prosodic component of intonation has potential for use in transgender communication therapy.

According to a recent literature review by Dacackis et al., findings on the contribution of voice quality to gender perception seem inconclusive, but there is evidence that a breathy voice quality contributes to a female sound as well.

**Speaking Fundamental Frequency and Vowel Formants as the Core Contributors to a Female Voice**

As any male speaker attempting to imitate a female will raise his pitch to a higher, lighter, tone, so too does the transgender MtF speaker strive to achieve a higher pitched voice. In fact, it appears fundamental frequency is the voice characteristic most frequently manipulated by transgender speakers on a conscious level as these male-to-female speakers raise their SFF to at least the boundary level, between 145 and 165 Hz. In general, the researched consensus is that elevated pitch appears to be the most salient cue for perception of a female voice in a biological male. Still, as we have discussed, it cannot explain gender perception on its own.

In 2009, Hillenbrand and Clark conducted a study to determine the contributions of fundamental frequency and formants in creating a distinction between perceived male and female voices. They did not include transgender people in their study, but the implications of their work are certainly applicable to the relevance of voice parameters in transgender voice therapy. The study utilized 25 sentences spoken by men, and 25 sentences spoken by women. Each sentence was made into 4 versions with the use of a source-filter synthesizer. Researchers left the sentence as it was for one version, and for the following three they (1) shifted the male voices up to a typical female $f_0$ and the female voices down to a male $f_0$, (2) shifted vowel formants only up to a typical female value for the male voice.
and down to a male value for female voices, and finally (3) shifted both the formants and the $f_0$ up or down respectively to values typical of the opposite gender of the original voice. From there, another set of normalized samples was devised so that formant frequency values could be presented in context as more typical of the specific vowel in which they presented. Researchers wanted to know if the vowel formants were high or low not in general, but in relation to the speech sound being uttered, which would reflect the (perhaps gendered) variations in the vocal tract size and how it affects the sound measurements. Listeners were asked to simply identify speaker sex for each of the manipulated and original sentences from both the absolute and normal sets for the first experiment. For the second, the listeners identified the speaker sex of isolated syllables containing a vowel sound, which was extracted from the longer samples.

The results of the study, as indicated in Table 1, concluded that SFF alone indicates speaker gender with nearly 96% accuracy in normalized samples, and that the three lowest formant frequencies of a vowel distinguish gender with accuracy at about 92% (see percent correct under normalized frequencies). Most profound, perhaps, were their discoveries that $f_0$ distinguishes the speaker’s sex more accurately alone than the formants do when considered alone, and that both of these features combined more accurately affect perception of the speaker’s gender (at about a 97% accuracy rate) than does either feature alone.\textsuperscript{6} When a single parameter was altered, listeners almost always perceived the gender of the original talker. Still, because formant frequency shifted voices of the opposite sex aren’t entirely convincing, these cues are not completely efficient in determining the speaker sex. For experiment 1, only 34% of pitch-shifted male samples and only 18% of
vowel formant altered samples were perceived as female voices. Similarly, 19.1% of pitch-shifted female samples and 11.7% of formant-shifted samples were perceived as male. Researchers also noticed that the major difference between experiment 1, involving identification of whole sentences, and experiment 2, identification of syllables, was that listeners were more resistant to the change in perceived speaker sex of a sentence than of a syllable. Perhaps then, they concluded, the other crucial residual effects lie in the variability of a speaker’s intonation patterns, or in the breathiness of a speaker’s vocal quality. Certainly then, we owe our inherently male and female sounds not only to our SFFs and vowel formants, but also to these extra cues.

In 2011, King et al conducted a study that actually involved MtF transgender patients, but studied some of the same parameters of voice and effect on gender perception in relation to whether or not MtF speakers were perceived as passing females. These researchers compared the voices of trans-people to those of biological females by having all participants from both...
groups read “The Rainbow Passage” as expressively as possible, and recordings were presented to 20 naïve listeners who judged them on a 7 point rating scale from extremely feminine sound to extremely masculine sound. Listener ratings grouped the speakers into M/M (male rated as male), M/F (male rated as female), F/F (female rated as female), M/M (male rated as male), as well as F/M (biological female rated as male). The main focus here was on the gendered perception and passing success of the MtF transgender clients, in hopes of fitting into the M/F group. Five speakers of the 30 MtF speakers were rated as female. With reference to Figures 1 and 2, above, the M/F and F/F groups showed consistently higher SFF and vowel formants than the M/M group. Still, strong statistical differences between the M/F and F/F groups were only shown on F2 of the vowel sound /i/ (p = .009), and on F3 of the vowel sound /ɛ/ (p = .012). Hence, the data suggests here that the differences in formants between biological females perceived female and biological males perceived female are not sufficiently high to yield any statistical difference that would suggest formants play a large role in listener perception of a MtF speaker as being female.¹

This conclusion does not directly contradict the studies of Hillenbrand and Clark by making any kind of suggestion that differences in vowel formants between males and females do not contribute to perceptual gender identification, but simply that vowel formants were not remarkably changed in MtF speakers that were able to switch their voice perception by a listener from male to female. These data simply support the salience of speaking fundamental frequency to this case, suggesting a low SFF limit of 140Hz and a mean SFF of 170Hz as the most likely contributing factors to perception of a voice as female. King et al do point to other studies which have shown that retracted lips and forward tongue carriage can manipulate resonance and elevate formants F2 and F3 to facilitate female voice perception in biological males. They also suggest (without evidence), that varying pitch and an exaggerated pattern of feminine intonation, or even a breathy vocal quality, may contribute as secondary cues to SFF to determining gender based on voice.

**Vowel Formants and Oral Resonance**

Why did so few of the MtF speakers in the aforementioned study pass as female? Why did the researchers argue that vowel formant frequencies were insignificant to perception of a transgender speaker as female while still acknowledging the usefulness of lip-spreading and forward tongue carriage as a means to increase perception of speakers as female? King et al. did not research the implications of manipulating oral resonance space and how this affects vowel formants, and consequently, gender perception. In 2013, Gelfer and Bennett sought the answers to this issue. They studied a selection of tall males, short females, and middle height males and females, and recorded them saying two different carrier phrases to two different vowel sounds in a sentence. They then took each sample (whether male or female) and digitally altered it to distinct SFFs that represent average female and male voices as well as the ambiguous voice range. Vowel formants stayed intact within 10% variation from the original measure. One could assume then, that for each sample that did not remain at the gendered SFF of the original speaker, the vowel formants were “mismatched” to the SFF. Listeners were asked to judge the gender of each of these speech samples.² Gelfer and Bennett acknowledged the boundary region of speaking $f_0$ around 165 Hz and the implication that other cues become more
important to gen-
dered perception if a
person’s speech lies within this area. Hence,
they studied vowel formant frequencies.
This study is unique because it chose
participants in height proportions “to
maximize potential gender based vowel
formant differences for two groups.”

Gelfer and Bennett report that a
study by Hillenbrand demonstrated that
because males have somewhat longer vocal
tracts, females tend to have higher vowel
formant frequencies. This is a property of
resonance and the effect of the space in
which the sound is created upon the acoustic
properties of vowel sounds. Large males
and small females were chosen for this study
so that there might be an extreme difference
in average vocal tract size between the two
groups and thus the resonance affecting
vowel formants.

The results indicated that the tall
male group was perceived as male reliably
at all but the highest SFF variation, and the
short females were perceived as female at all
SFF variations. This is likely due to
gendered characteristics of vocal tract size
that contribute to resonance and the
respective prominent sound quality created
by the vowel formants. Middle height
speakers were consistently perceived as the
correct gender as well: females at all SFFs,
and males with slightly less consistency in
the two highest SFFs. Thus, changing SFF
to a value characteristic of the opposite
gender of the original sample was only
sometimes successful in changing the
listener’s perception of the actual sex of the
speaker. When all results across groups for
males are averaged, 49.3% were still
identified as male when the SFF was shifted
to female levels. Females on the other hand
were perceived female 85% of the time,
even when SFF was shifted to male levels.
Listeners perceived the gender of women
correctly most of the time, and a little less
than half of the time for men. Thus, the
study found that vowel formants do indeed
contribute to gender perception, especially
in the SFF boundary region. Even when
SFFs were mismatched with gender-typical
vowel formants, the distinctive sounds that
the formants created led to accurate
identifications of the speaker’s original sex.

Findings that support biological
gender differences as shaping the perception
of a voice as male or female may come as
alarming news to transgender clients and
their speech therapists. However, research
by Carew et al can reassure the effectiveness
of oral resonance therapy for MtF
transgender speakers as a means of attaining
these inherently female vowel formants.
These researchers cite evidence that
changing the shape of the vocal tract at the
mouth through lip spreading and
constricting the vocal tract by means of
anterior tongue placement can change the
formant frequencies that speakers produce in
vowels. They studied ten male-to-female
clients, taking acoustic measures pre-
therapy and after five sessions of oral resonance
vocal therapy, both of the clients reading
“The Rainbow Passage.” Recordings were
also presented to listeners who rated the
femininity or masculinity on a scale.

Results of this study indicated that
all three vowel formats (F1, F2, F3) of all
three vowels analyzed increased post-
therapy according to the group mean, and
four of the participants were perceived
increasingly as feminine by listeners.
Overall trends suggest forward tongue
 carriage and lip spreading as successful
techniques for increasing vowel formant
frequencies, which in turn do appear to
increase listener perceptions of femininity of
voice. Even if one of the most influential
vocal parameters for gender perception is
based on the shape of the anatomy of the
source, there are successful methods for
maneuvering around this physical obstacle,
as indicated here by Carew et al.
Breathiness and Intonation as Cues for Gender Identification

Nearly all of the studies of fundamental frequency and formants suggest that while they may be the most salient vocal parameters for listener identification of gender, they cannot possibly stand alone as sole indicators of voice gender, especially in Wolfe’s boundary region as cited by Gelfer and Bennett. Hillenbrand’s study concluded with a suggestion of the importance of other cues such as prosody and breathiness. Similarly, King et al propose that varying pitch and a pattern of feminine intonation, as well as breathy vocal quality, may contribute as secondary cues to SFF.

Several studies reference the perceived voice quality of breathiness as a possible secondary cue to our essential parameters of fundamental frequency and vowel formants, but there is very little empirical evidence that breathiness does in fact have this effect. In 2007, VanBorsel et al. conducted a two-part experiment to determine the extent to which breathiness contributes to listeners’ perceptions of femininity. Seven biological female speakers were recorded producing a normal “ahh” vowel and a breathy “ahh.” In the first experiment, all recordings were presented to listeners as if they were samples from 14 different speakers, and they were prompted to rate femininity on a five-point scale from “little feminine” to “very feminine.” In the next, normal and breathy recordings from each speaker were presented as a pair and the listeners were required to determine which of the two sounded more feminine.

Results of experiment 1 showed that breathy samples always received higher femininity scores than their natural sample. Experiment 2 was slightly less reliable but still results indicated that more listeners judged the breathy sample to be more feminine than the normal component. Certainly one may draw the conclusion that breathiness contributes to perceptions of the voice as being feminine. Still, it is important to realize that this study involved only female speakers and did not study the effects of breathiness upon listener perceptions of MtF voices. Thus, VanBorsel et al advise caution when generalizing these findings to voice therapy with male to female transsexuals.

Measurements of aerodynamics and acoustics do not necessarily correlate with the perception of breathy voice quality by listeners, and thus to determine the relevance of these measurements to gender identification, it may be helpful to include groups with differences in perceived voice quality.

In addition to breathiness, another parameter that much of the research suggests should be considered as contributing to femininity of voice is intonation patterns. Like breathiness, this parameter of voice has not been extensively researched with respect to its implications for transgender speakers and how it can be used to enhance feminine perception. Studies in 2014 by Hancock et al. attempted to address this shortcoming of the research by examining the intonation of biologically female, biologically male, female-to-male, and male-to-female speakers describing a Norman Rockwell image. They compared measures of intonation between biological gender groups, perceived gender groups, and between MtF speakers who were perceived as male, female, or ambiguous.

Participant speakers were recorded talking about the image for about 30 seconds, and listener participants were presented the recordings and asked to rate the voices they heard on a perceptual scale of masculine male on the far left to feminine female on the far right with feminine male/masculine female in the center of the scale. Acoustically, the recordings were
analyzed for intonation shifts and divided into upward and downward intonation shifts. All female speakers were perceived as female, and all male speakers were perceived as their biological sex as well. 4 of 14 MtF speakers were perceived as female and 5 of the 6 FtM. The group that was perceived by listeners as female used a greater pitch shift range and a greater number of upward intonations, and among MtF speakers, those passing as female used more upward intonations and fewer downward intonations. Although no technical statistical difference was appreciated for any intonation measure, and it cannot be determined that intonation is a clear indicator of whether or not a transgender person is perceived as female, the data show that a high percentage of utterances with downward intonation may cause a speaker to be perceived as male. The researchers suggest that the role of intonation in gender perception appears to be minor, but is not completely irrelevant.

Implications for Further Research

Despite significant evidence that $f_0$ and vowel formants are the most salient cues to a listener’s judgment of a voice as male or female, nearly all studies suggest that these parameters alone are not responsible for a person’s gendered voice identity. These studies point to breathiness and intonation as possible other cues to consider as significant to gender perception. Unfortunately, both parameters lack significant empirical evidence as substantial contributors.

VanBorsel et al found that voice samples with a breathier quality are more often perceived as feminine than are regularly spoken voice samples, but the researchers admit a need for further investigation into what specific degree of breathiness constitutes a passing feminine voice. These scientists acknowledge the shortcoming of their research in that all speakers studied were female, perhaps implying (in application to a MtF speaker’s situation), that for a client who had already achieved an acceptable female $f_0$, breathiness could further improve perception of the voice as belonging to a female. As physical applications for breathiness on the perception of a male or a transgender MtF voice have not even been studied, certainly even a similarly structured study involving samples from these populations would assist clinicians as they assess the usefulness of breathiness training as part of transgender voice therapy session.

Another obvious shortcoming of the breathiness study at hand is that while we have clear numerical values for fundamental frequency levels that are suitable for facilitating feminine voice perception, there is not an actual identified measure of the degree of breathiness necessary to facilitate a more feminine voice production. In this study, the women speakers were simply instructed to speak in a breathy voice, and listeners utilized a subjective scaling measure to rate the breathiness level on a scale of 0-3. Although we are aware that breathiness indicates some level of increased spectral noise, according to Dacackis et al. it is understood that the voice quality listeners hear is not necessarily correlated with the acoustic and aerodynamic measurements that are collected. Future research must reach a means by which to numerically target the physical increase in spectral noise that is characteristic of perceived “breathiness.” It seems, according to Wayland et al and as cited by VanBorsel et al, that it has been difficult for researchers to create a standard “computational means” by which to distinguish breathy voicing from clear voicing. The research determines that the next step toward finding the breathiness cutoff value for femininity will be the creation of a more concrete numerical scale.
to measure the level of spectral noise in a sample incrementally. Then, studies can be conducted to determine how this breathiness cutoff effects perceived femininity of biological female, biological male, and male-to-female speakers, and whether or not the cutoff value is the same or different for each of these groups. Ultimately, it can then be determined how effective the achievement of this feminine breathiness value is in allowing a MtF speaker to pass as female.

Researchers find shortcomings in studies of intonation impact on the perceived femininity of voice as well. Though Hancock et al were unable to prove with statistically significant numbers that intonation helps MtF speakers pass as female, it was obvious from the data collected that voices perceived as female contained a greater number of upward utterances within a greater pitch shift range, and those perceived as male generally show a pattern with more downward intonation. These findings do then support the implementation of intonation measures in the therapy setting as a minor addition to the work of a MtF client attempting to pass as a female to listeners.

Still, much of this data is purely perceptual. Such conclusions beg the question: Can we find a measurement of intonation patterns that ensure female voice perception? Hancock et al agree standardized measures of intonation should be established in order find target components of overall intonation that will allow the client to achieve this added feminine quality when combined with the necessary adjustments in SFF and vowel formants. The conclusions even consider that further research should investigate the assumption that working on feminine intonation patterns may simultaneously facilitate the necessary gender-related changes in $f_0$. This can be accomplished by studying the acoustic evolution of $f_0$ over the course of intonation-based therapy for MtF speakers from the baseline until the optimum level of “female” intonation is reached and comparing baseline and discharge values of $f_0$. Once standard measures of intonation are identified, researchers could easily study changes in fundamental frequency as uniform intonation-feminizing techniques are implemented in transgender voice therapy.

It is clear from the concluding remarks of studies conducted thus far that further research on this topic is necessary to truly determine which vocal parameters must be targeted in therapy to achieve optimal results for male-to-female clients attempting to pass as female while communicating. Studies that draw conclusions about the salience of fundamental frequency and vowel formants suggest looking further at breathiness and intonation. Though there is preliminary evidence of the importance of these to a passing female voice, researchers agree that standardized methods for measuring breathiness and intonation will clarify the specific degrees to which these voice parameters must be altered to effectively assist in gendered voice perception. Then, studies may take a closer look at therapy techniques that will best aid the client in achieving target measures for feminine breathiness and intonation.

**Conclusions**

The studies of Hillenbrand and Clark as well as King et al reaffirm the previously held notion that speaking fundamental frequency is the most salient vocal cue to gender perception, while also acknowledging the addition of increased vowel formant frequencies as crucial to the perception of a speaker as female. Though King et al indicates a lack of statistical evidence that vowel formant
frequencies are essential to the MtF transgender speaker’s goals to pass as female, these researchers do acknowledge the correlation of increased vowel formants F2 and F3 to the perception of femininity of voice. Their suggestion that these formants can be increased through forward tongue carriage and lip spreading is supported by the findings of Gelfer and Bennett, which indicate that the biological size differences between the male and female vocal tract play a role in the identification of voice as male or female as a result of different resonances that produce lower and higher vowel formants, respectively. The work of Carew et al reassures us that transgender clients can attain similar resonance results to those inherent of the female vocal tract through therapy techniques that stress this anterior tongue placement and manipulation of the tract at the mouth by means of lip spreading. Their results suggest that MtF clients were more likely to be perceived as females by judges following a course of oral resonance voice therapy.

Nearly all of these studies agree that resonance and fundamental frequency addressed alone in therapy cannot guarantee successful achievement of the feminine voice sound. Thus, there has been preliminary research on the contributions of breathy vocal quality and intonation as cues to speaker femininity. Studies have concluded that there are implications for the effectiveness of both when addressed in therapy, but the lack of empirical evidence and quantitative targets for these parameters make them harder to address in a uniform manner. For now the findings support fundamental frequency and vowel formants increased in combination as the most salient voice parameters to the perception of a speaker as female. As breathiness and intonation are studied further, perhaps they may take on more prominence in the process of voice change from male to female. For the time being, the most important parameters of voice to focus on in therapy are increased speaking fundamental frequency and resonance, as they have been empirically proven to improve instances of identification of a speaking voice as female.

References


