A Longitudinal Acoustic Study of Two Transgender Women on YouTube

Andrew Cheng, UC Berkeley

1 Introduction

Following a global rise in the visibility of people who identify as transgender or non-binary around the world (Zabus and Coad, 2014), interest in the acoustic characteristics of transgender peoples' voices has greatly increased. Most academic studies of transgender voices come from the field of speech pathology, with papers from journals in communication disorders, laryngology, and the like, describing best practices for treating individuals who wish to transition, or comparing transgender voices to cisgender voices on a number of scales (see Davies et al. (2015) for a recent overview). Though these studies are useful in a clinical setting, they often implicitly put on a pedestal only one way to achieve the ideal masculine or feminine voice. The practice of speech pathologists who work with transgender clients are, of course, based in decades of linguistics research that has deeply explored all of the acoustic correlates of perceived gender (Simpson, 2009; Cameron and Kulick, 2003), and it is clearly beneficial to transgender individuals whose quality of life often depends on their ability to be perceived as sounding like their identified gender. Still, when the overwhelming majority of linguistic studies on transgender speech focus on the means to an end, the field risks compromising a scientific commitment to document variation as it exists by focusing too narrowly on clinical cases, as well as contributing to a kind of normativity for gendered voices, the same kind that has shadowed all studies of gender ever since it was shown to be a socially constructed phenomenon (see Eckert and McConnell-Ginnet (2013)).

The current study addresses this normativity in two ways. First, it is a study of transgender voices outside of the clinical setting: voices that belong to transgender individuals who desire to change how their voices are perceived, but are not undergoing direct treatment or medical intervention of any kind to do so. Second, it tracks their vocal characteristics over many years and finds that not only are their voices following completely different trajectories as time progresses, they are in several ways deviating from the expectations for their gender. Obviously, if a transgender individual does not follow a particular treatment program, their voice is unlikely to change in the way the treatment program would predict. However, this doesn't mean that the individuals are any less successful in their transition. The study concludes by speculating about the myriad ways in which a transgender person may use vocal and visual cues to index their gender, despite not changing their voice in the specific, most salient ways one might expect, given past linguistic research on gender.

It is important to many individuals who transition from one gender to another to adopt speech patterns that allow them to pass in society as male or female. Female-to-male transgender individuals (FTM, or trans men) may elect to undergo testosterone treatments. This lengthens their vocal folds, which consequently lowers their pitch, or fundamental frequency (f0). Male-to-female transgender individuals (MTF, or trans women) often undergo hormone replacement therapy with estrogen in order to feminize their body, but estrogen cannot shorten or decrease the thickness of the vocal folds and raise the vocal pitch of any person. It is thus commonly thought that trans women cannot undergo any medical treatment aside from invasive surgery (phonosurgery, cricothyroid approximation, laryngoplasty, etc.) in order to permanently raise the fundamental frequency of their voices, which is not recommended by all medical professionals (Davies et al., 2015). (It would be wise to note here that whether or not a transgender individual undergoes hormone replacement therapy or gender confirmation surgery does not affect how trans they are; a transgender individual, as with all human beings, is simply the gender they identify as, regardless of their body.)

However, it is understood that there is wide variability in the "average" pitch ranges for men and women. Not only are there plenty of men with higher-pitched voices than many women and vice versa, even within an individual, drastic pitch changes can occur in speech, depending on the situation. In addition, pitch is just one of many acoustic markers of gender identity. There are many other variables besides fundamental frequency that speakers use to index their gender (or that listeners use to cue in on speaker gender), such as f0 variation and prosody, sibilant frequency, and vowel quality, which may be less salient of a cue than pitch.

Because of this, a MTF transgender individual can, without clinical treatment, successfully attain a female-sounding voice through practiced alternation of fundamental frequency and an increase in use of other sociophonetic variables that index femininity. However, the current study is less concerned with a hypothesis about what kind of voice a trans woman is capable of attaining, and focuses more on the kinds of variation that exist in her voice during and after a transition. Indeed, findings suggest that a trans woman can use different combinations of all the sociophonetic variables that index femininity, at different times, and thus over time her use of any one specific variable may not necessarily increase or decrease in the expected direction.

2 A few ethical points

This study examines several acoustic variables in the speech of two trans women who have kept up regular video diaries on YouTube (‘vlogs’) for many years. Before going into the methodology, I believe it is important to note here for the informed reader that conducting research on members of a protected minority requires a level of respect and care beyond the standards for scientific research with human subjects. Readers may be concerned about the use of data taken from the Internet without informed consent and/or about the positionality of the researcher with respect to the subjects and to the transgender community overall, and his intentions behind the research.

To address the first concern, it is generally accepted that content taken from publicly available sources such as YouTube (i.e., any person with Internet access can freely access all content) is fair for use as data in research without seeking informed consent from its creators. The nature of YouTube vlogs, and of the trans vlogs in particular, is meant in part to broadcast the vlogger’s lives and experiences to as large an audience as possible, and the vloggers are highly aware that all the content they produce will be watched, scrutinized, commented on, and perhaps even scientifically analyzed.

At the same time, it is true that research with minoritized subjects has often had unintended negative social consequences for the subjects. Although transgenderism is increasingly visible in the public sphere, it remains a highly stigmatized category and, even when it is accepted, it can still be ‘Othered’ (exceptionalized, infantilized, exoticized, and generally misrepresented due to being non-normative) by commentators. There is indeed a history of exploitative research on trans and other queer subjects, and I seek to avoid that, especially from my positionality as a cisgender (non-transgender) scientist.

I have read and taken to heart the discussion of the ethics of studying transgender vloggers on YouTube in Raun (2010), and I have chosen not to anonymize the two women who are the focus of this study. However, I refer to them by their YouTube usernames and not their real names. I have also left out any information about their current hometowns. I do this with the intent to give readers the opportunity to visit the vloggers' channels and to hear them tell their own stories in their own voices.

To address the second concern, I acknowledge that I am not an insider in the trans community, as I identify as cisgender. Thus, it is reasonable to cast suspicion on my motives as a researcher, to believe that I may be using the trans community as just a tool to explore some theoretical notion or supplement a research agenda that does not benefit the trans community. Indeed, I began this project as part of an exploration of non-binary identities for a class I was teaching on language, gender, and sexuality. Using Davis et al. (2014) as a jumping off point, I wanted to show my students that the long-held notions of gender binaries were only social constructions and could be subverted through language and the voice, and that transgender individuals were a good example of such.

But the more I listened and read, the more I realized that I wanted my research to do more than just prove a theoretical point. Although this is a study on the acoustics of two people's voices and not an analysis of the trans narrative as a literary style or trans as an object in cultural studies, I hope that this research may grow the visibility of trans identity in this academic field and to do so as respectfully as possible. I hope that this project is understood to center my subjects and their voices, not any particular hypothesis or theoretical approach.

Lastly, I take seriously Raun's point that "within feminist and activist knowledge production, as well as within the tradition of transgender studies, research is not separated from but grows out of everyday practices and politics, and involves dialogue with the people involved" (Raun, 2014, 27). To this end, I have informed both Grishno and PrincessJoules that I am carrying out this research. Grishno has affirmed the project and consented to work, although PrincessJoules has not replied to my communications. (It is for this reason that I have chosen to refer to both of them by their YouTube usernames.)

3 Methods

3.1 Speakers

The data, as mentioned above, come from two trans YouTube vloggers. The first is a White American activist and student who goes by Grishno on YouTube. She began making videos about her life as a transgender woman in November 2006 and has since uploaded nearly four hundred videos (an average of two or three videos a month for 140 months). Most of her videos employ the typical 'talking head' YouTube vlog format of the speaker sitting in front of their camera, centered, and narrating about their life or a specific topic, with few cuts or other edits such as music or animations (Horak, 2014). As one of the very first trans vloggers on YouTube (which was launched in February 2005), Grishno is considered to be a pioneer of the art. She has produced a documentary about her trans life and has been interviewed for several other academic studies of transgender YouTube vloggers.

The second vlogger is a Vietnamese Canadian model and makeup artist whose YouTube handle is PrincessJoules. Since 2011, she has recorded over five hundred videos (an average of six or seven videos a month for 80 months), which also use the talking head format, though occasionally with music and sound effects added in. PrincessJoules has amassed hundreds of thousands of subscribers in a relatively short amount of time due to the frequency of her uploads as well as the subject matter: PrincessJoules' videos are most often on the topic of her transition process (with detailed stories and photos of transitional surgeries, as well as how-to advice for other MTF individuals) or beauty and makeup tutorials. She has been interviewed in local media for her activism and prominence in social media.

Both women have undergone gender confirmation surgery and hormone replacement therapy, but neither has undergone vocal feminization surgery.

3.2 Corpus

Data were taken from twenty-seven of Grishno's vlogs and twenty-three of PrincessJoules' vlogs, for a total of fifty videos and a little over four hours of speech. Grishno's vlogs ranged from one to ten minutes long (average duration: 4:22) and were uploaded between January 2007 and December 2017. PrincessJoules' vlogs ranged from three to ten minutes long (average duration: 6:17) and were uploaded between September 2011 and January

2018.

3.3 Analysis

Every video was transcribed in Praat (Boersma and Weenink, 2018) and force-aligned (Yuan and Liberman, 2008; Rosenfelder et al., 2011). Vowel formants, vowel duration, and fundamental frequency were measured using the IFC formant tracker (Ueda et al., 2007) and related tools developed through the Berkeley Phonetics Machine (Sprouse and Johnson, 2016). The tracker extracts values in Hertz directly from the sound wave every five milliseconds. These values were merged with the force-aligned data so that each segment would have a number of values relative to its duration (e.g., a 30-millisecond vowel would have approximately six f0 measurements. Frequency-based measurements such as f0, F1, and F2 were not normalized for speaker, since even though the two speakers have different voices and accents, all statistical analyses were run on each speaker separately. Vowel duration for each speaker was normalized for local speech rate by dividing the duration of each vowel by the number of syllables in a rolling thirty-second window.

In addition, every video was tagged with metadata concerning the chronological relationship of the video to the speaker's transition (e.g., pre-transition, post-operation, etc.), as well as the video topic and keywords. Topics included general categories for videos, such as About My Day, Political, or Rant. Keywords were more specific and included terms such as *marriage*, *surgery*, *passing*, *college*, and *fashion*, with several keywords per video.

The sociophonetic variables that are the focus of this study are fundamental frequency, F1 and F2 (vowel formants), vowel duration, and sibilant duration. For each variable, the average value for every video was calculated, and then each video was treated as one token in a longitudinal analysis. Change over time is measured by change in these averages from one video to the next. In addition, each video was divided into thirty equal-length bins, such that a two-minute video would have thirty 4-second bins and a 10-minute video would have thirty 20-second bins. The trajectory of each of the variables throughout the duration of a video could then be analyzed as change over video duration.

4 Results

4.1 Change over time

4.1.1 Fundamental frequency (f0)

Fundamental frequency (f0) is determined by the frequency of vibration of the vocal folds in the larynx. To reiterate, no hormone therapy can cause the vocal folds to shorten or become thinner, but fundamental frequency can be changed by raising or lowering the larynx, which tightens or slackens the vocal folds.

A simple linear regression was calculated to predict f0 over chronological time for each speaker. Significant regressions were found for both Grishno (F(1,25)=25.917, $p<0.001$), with an adjusted R^2 of 0.489, and PrincessJoules (F(1,21)=7.0826, $p=0.015$), with an R^2 of 0.217.

As shown in Figure 1, Grishno's mean fundamental frequency decreased as the years progressed, going from well above 150 Hz in the earliest videos from 2007 to around 125 Hz in the most recent videos in 2017. PrincessJoule's fundamental frequency started out above 150 Hz in her first videos, even in the first two videos in late 2011, before she came out as transgender, and has steadily increased over time. Figure 1 has also been labeled with pertinent milestones in each speaker's lives, regarding their transition or their YouTube career. Grishno's timeline includes her first video upload to YouTube, her orchiectomy, her relocation to California, and her two gender confirmation surgeries. PrincessJoules' timeline includes her coming out video, her gender confirmation surgery, and her breast augmentation surgery.

A fundamental frequency of 150 Hz does fall within the generally-accepted range of fundamental frequencies of cisgender male speakers (Simpson, 2009), or a bit lower than the lower end of the range for cisgender female speakers. From that point, an increase or decrease of 25 Hz is not that much, considering the wide range of possible frequencies in the human voice. But it is likely discernible. At least in pure tones, an increase from 150 Hz to 175 Hz is approximately three semitones (or a minor third interval).

Immediately, what this reveals is that the two trans women's voices changed in opposite directions over the course of their transitions, and afterward. If a relatively higher pitch is an index of feminine speech, then PrincessJoules' voice has become slightly more feminine over time, while Grishno's voice has become less masculine. But of course, pitch is not the sole

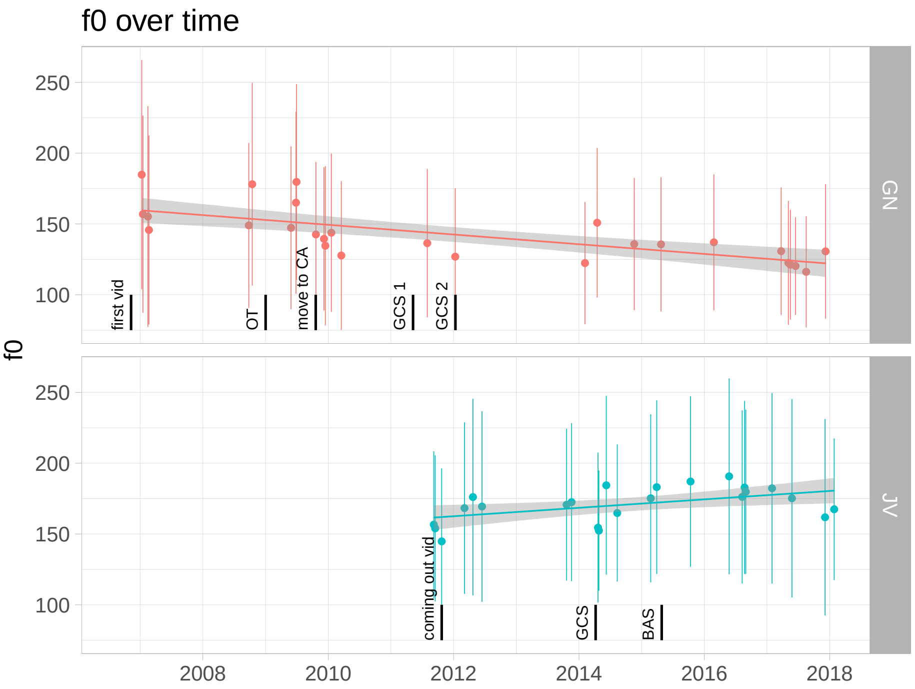

Figure 1: Mean f0 for each speaker is plotted, along with error bars representing one standard deviation above and below the main. Mean f0 for Grishno (GN) decreased significantly over time, while mean f0 for PrincessJoules (JV) increased significantly over time. The black bars represent significant events in the speaker's YouTube career or in their gender transition.

determiner of what makes a voice feminine or not, and so we turn to the other variables.

4.1.2 Vowel formants

The measurements for F1 and F2, which determine vowel quality, were averaged for each video. The values for formants comes from the 'filter' of the voice, or the vocal tract. An individual may slightly lengthen their vocal tract, which has the effect of lowering all formant values. But formant values for vowels are also affected by the shape of the tongue and lips.

Simple linear regressions were calculated to predict F1 and F2 over chronological time. Regarding F1, a significant regression was found for Grishno

(F(1,25)=4.3049, p=0.048), with an R^2 of 0.113. Regarding F2, significant regressions were found for both Grishno (F(1,25)=27.181, p<0.001), with an R^2 of 0.502, and PrincessJoules (F(1,21)=13.926, p=0.001), with an R^2 of 0.37. Overall, Grishno's F1 has decreased over time. In general, men have been shown to have lower F1 and F2 values than women (Simpson, 2009), but this is mostly considered to be due to the longer vocal tract length of men. Since Grishno has not undergone any procedures to lengthen her vocal tract, the significant change over time for her formant values must be due to difference in articulation. PrincessJoules, on the other hand, has a significant increase in F2, although this increase is only part of the full picture of her vowel articulations. For the most part, the directions of change for both speakers are identical to those for their f0, but a more detailed analysis, broken down by vowel, is necessary to interpret the changes meaningfully.

To that end, a simple linear regression was calculated to predict F2 of back vowels over chronological time: specifically, /u/ as in *two* and /oʊ/ as in *toe*. For PrincessJoules, a significant regression was found for F2 of /u/ only (F(1,21)=10.418, p=0.004), with an R^2 value of 0.3, indicating slight backing of /u/ over time. For Grishno, the vowel /oʊ/ decreased in F1, judging by a significant regression (F(1,25)=5.317, p=0.03) with an R^2 of 0.142, as well as F2, judging by a significant regression (F(1,25)=22.474, p<0.001) with an R^2 of 0.452. This indicates that Grishno's /oʊ/ backed and raised over time, while /u/ did not change significantly. For PrincessJoules, there was little change in back vowels apart from slight backing of /u/ over time.

For front vowels, such as /ɛ/ as in *ten*, /i/ as in *tee*, and /æ/ as in *tan*, the same linear regressions were performed. For /æ/, the low front vowel, a significant linear regression on F1 was found for PrincessJoules (F(1,21)=8.575, p=0.008) only, indicating that her /æ/ utterances have lowered (indicated by an increase in F1) over time. On the other hand, a significant linear regression on F2 was found for Grishno (F(1,25)=10.896, p=0.003) only, indicating that her /æ/ utterances have backed (indicated by a decrease in F2) over time. Regarding /ɛ/, a significant regression was found for both Grishno (F(1,25)=17.033, p<0.001) and PrincessJoules (F(1,21)=10.463, p=0.004). Grishno's F1 decreased significantly over time, while PrincessJoules' increased.

Table 1 combines the results of all the regressions, showing which vowels changed in F1 and F2 or did not change over time. Within an individual speaker, an increase in F1 indicates an articulation that is lower in the mouth, while a decrease in F1 indicates a higher vowel. An increase in F2

	vowel	F1	p-value	F2	p-value
GN	/i/	-	0.025	0	
	/ɛ/	-	<0.001	0	
	/æ/	0		-	0.003
	/oʊ/	-	0.03	-	<0.001
	/u/	0		0	
JV	/i/	0		+	0.002
	/ɛ/	+	0.004	+	<0.001
	/æ/	+	0.008	0	
	/oʊ/	0		0	
	/u/	0		-	0.004

Table 1: Change in F1 and F2 for various vowels for each speaker is indicated by + for an increase in Hz, - for a decrease, and 0 for no significant change.

follows a vowel produced more toward the front of the mouth, and a decrease corresponds to an articulation that is further back.

It is easier to visualize the change by plotting the F1 and F2 values on a vowel chart. Figure 2 below was created from the models for each vowel, where t1 indicates a point on the regression line that corresponds to an early date in their YouTube career, and t2 indicates the present day. The solid shape thus approximates the past vowel space, while the dashed shape indicates where the vowels have ended up today.

According to the figure, Grishno's vowel space has, as a whole, moved up and back over time, while PrincessJoules' vowel space has expanded evenly outward. Whether women generally have wider vowel spaces on average than men has been claimed but also contested (Simpson, 2009).

4.1.3 Segment duration

Raw segment duration was calculated from the automatic alignment. These values were then normalized for local speech rate by dividing each raw value by the number of syllables in a thirty second window surrounding each segment. Simple regression models were calculated to predict average vowel duration and average sibilant duration (for /s/ and /z/) over chronological time. It was found that both speakers' average /s/ and /z/ duration decreased over time, but the regressions were only slightly significant (p=0.093, and p=0.092, respectively). Vowel duration did not change significantly for

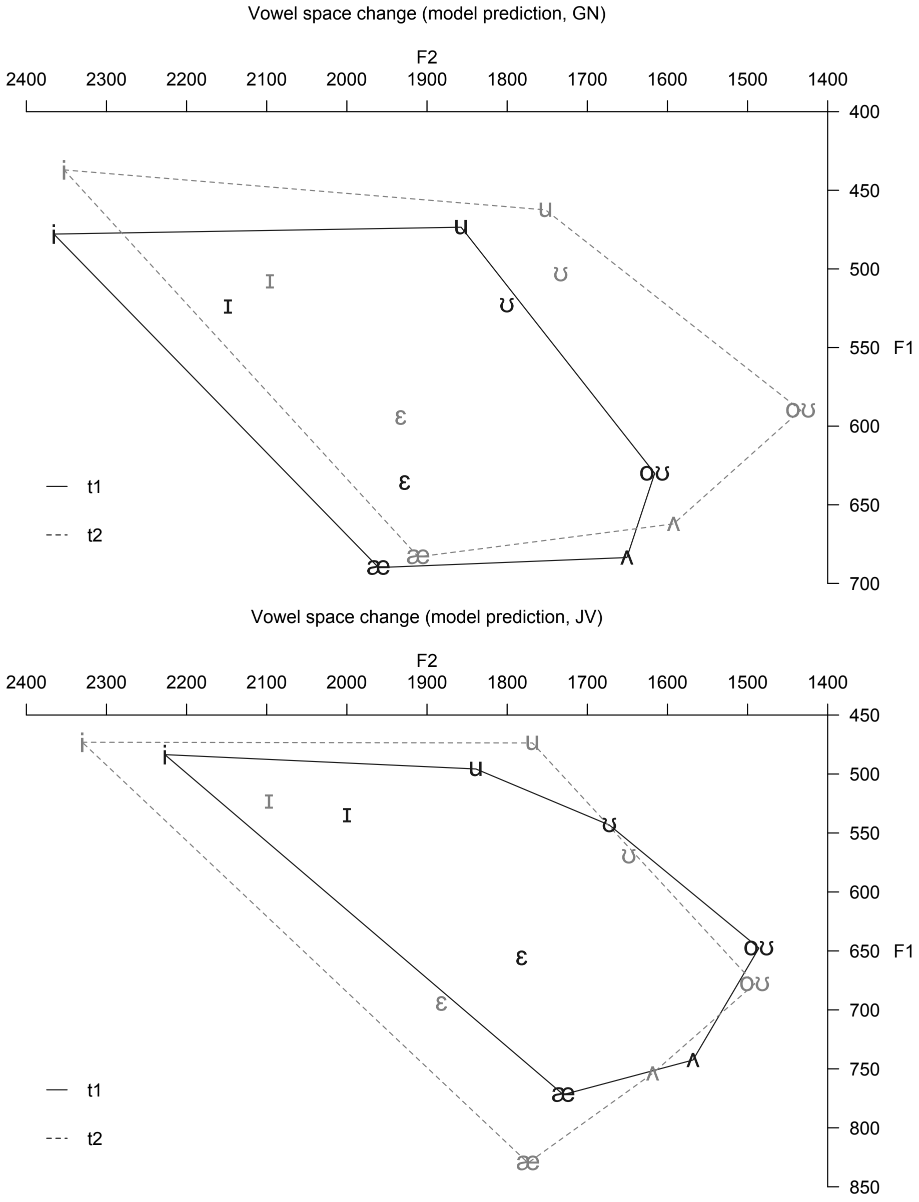

Figure 2: The linear regression models of each speaker's vowel space was used to predict F1 and F2 values for each speaker at the beginning of their YouTube career and the present day. These are plotted, with the beginning as t1 and the present day as t2.

either speaker over chronological time.

4.2 Change over video duration

New media scholarship has only just begun to scratch the surface of the YouTube vlogging platform, especially within the realms of sociolinguistics and conversation analysis. Horak (2014) explains that the talking head style that is ubiquitous among amateur vloggers, including trans vloggers, is unique to the platform in that it mimics a face-to-face conversation between the vlogger and their viewers, even though it is actually "nonsynchronous, unidirectional, and one-on-many" (Horak, 2014, p. 575). A single video is a monologue that is intended to be part of a dialogue, even though the interlocutors are removed in both space and time, and their responses are more often written (as viewer comments) than spoken (through response videos).

Frobenius (2014) has studied vloggers' strategic use of language to construct roles for their viewers. Building on Bell's framework of audience design (Bell, 1984), Frobenius argues that individuals who give monologues (on YouTube but also in traditional formats such as radio or television) may not expect the audience to interrupt (Clark, 1996), but their speech does "[contain] traces of interaction... linked to audience design" (Frobenius, 2014, p. 60). Viewers of YouTube vlogs may be considered addresses, auditors, or overhearers; whether or not the vlogger considers the viewer to be more of a participant (addressee or auditor) or not (overhearer) influences their lexical choice. For example, vlogs addressed to viewers that are meant to participate more will use generalized terms of address (e.g., "you guys") as well as imperatives. In addition, a vlogger may modulate the volume of their speaking voice to indicate shifts in framework (e.g., addressing people out of the camera frame in real time with a louder voice, then returning gaze and address to the camera at a normal volume). In these ways, audience involvement, though entirely dependent on how the vlogger chooses to frame their imagined audience, is crucial to the speech styles used in a YouTube monologue. For that reason, it is hypothesized that the vlogger's direct awareness of the ratified participant viewer will influence their use of pitch and other sociophonetic variables throughout the course of a video.

Recall that these simple talking head videos employ little to no cuts or other types of editing. Thus, the final monologue as presented in the video is mostly uninterrupted speech. Over the duration of a video, it is expected that the vlogger will pay more attention to their imagined interlocutors (the viewers) at the beginning and end of a video, with less attention in the middle. This will take the form of greetings and sign-offs and, in general, more of a performance, compared to points in a video when the vlogger will pay less attention to her speech. The greater attention should roughly correlate to increased (subconscious) use of sociophonetic variables that index gender and other identities, such as “YouTube personality” or “female”.

4.2.1 Fundamental frequency

As mentioned previously, each video was divided into thirty equal-length bins. Then, a quadratic model was calculated to predict f0 over video duration for each speaker. This meant that over the duration of a single video, the speaker’s f0 may follow a curved pattern, either decreasing and then increasing again, increasing and then decreasing again, or any other kind of change that is not strictly linear.

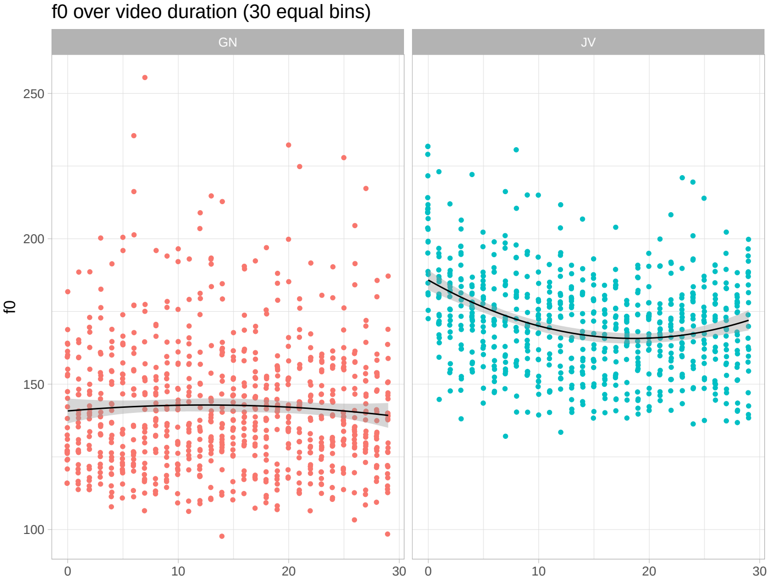

Figure 3: Mean f0, for each of thirty equal-sized bins of all videos, is plotted (color-coded for speaker), along with quadratic regression lines.

A significant regression was found for PrincessJoules only (F(2, 686)=40.579, $p<0.001$), with an R^2 value of 0.103, which is somewhat low[1]. As shown in

[1] Thanks to Keith Johnson for pointing out that the model is attempting to find the

Figure 3, PrincessJoules' videos often began and ended with speech at a higher fundamental frequency, while the middle of her videos saw the f0 dip.

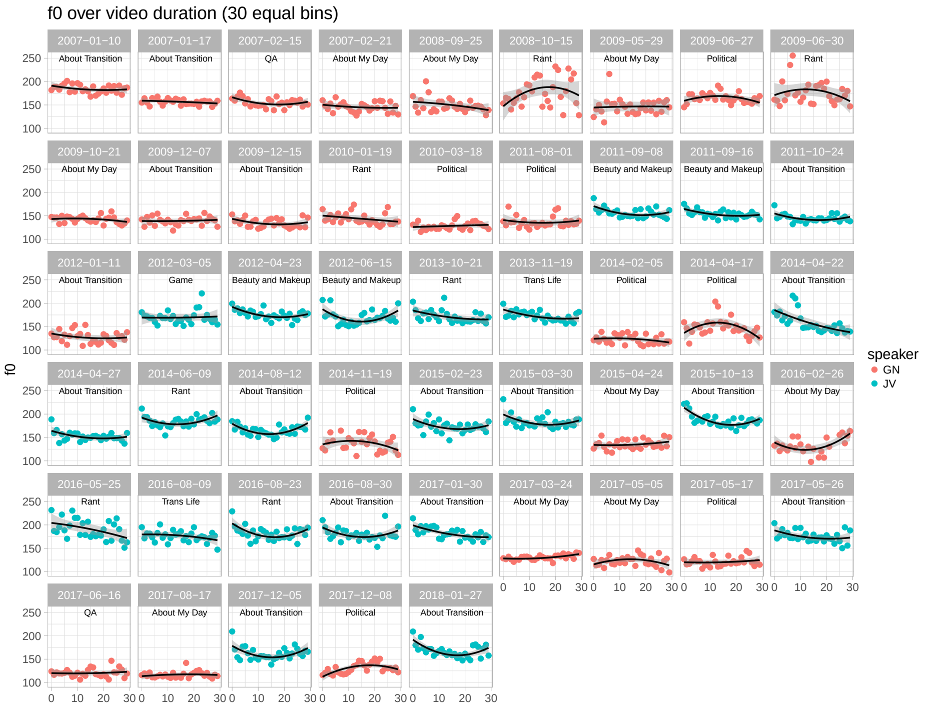

Figure 4: Mean f0, for each of thirty equal-sized bins of each video, is plotted (color-coded for speaker and labeled by video category), along with quadratic regression lines.

This is broken down by video in Figure 4, which reveals how individual videos may differ. In particular, it is noted that Grishno's videos tended to have a more monotonic f0, with little change across bins, with the exception of her videos tagged as Rant and some tagged as Political. For these videos, the middle sections tend to have higher mean f0 compared to the beginnings and ends, which is a reasonable finding in accordance with the fact that higher f0 is correlated with expressions of the emotions of stress and anger (Johnstone and Scherer, 2000).

The same trend of higher frequencies at the beginning and end of videos

same curved pattern for every video, rather than a unique curve for each video, which likely accounts for this low value.

was also found for PrincessJoules' vowel formants: significant regressions for F1 (F(2,686)=7.875, $p<0.001$), with an R^2 value of 0.02, and F2 (F(2,686)=5.101, p=0.006), with an R^2 value of 0.012. A quadratic model run on Grishno's F1 also returned a significant result: F(2,800)=3.545, p=0.029, with an R^2 value of 0.006. However, it is difficult to interpret these statistics, since the R^2 values are very low, which means that the model, while accurate, does not account for a high percentage of the patterns we see in the data.

4.2.2 Vowel duration

For vowel duration, significant regressions were found for both Grishno (F(2,800)=4.178, p=0.016), with an R^2 value of 0.008 and PrincessJoules (F(2,686)=3.625, p=0.027), with an R^2 value of 0.008. The patterns for each speaker follow some sort of curve (rather than being linear or unchanging), but, just as with the F1 and F2 results, it is not always the same kind of curve. Hence, the low R^2 value, indicating that the model does not account for most of the variation in the data.

As shown in Figure 5, both Grishno (GN) and PrincessJoules (JV) tend to begin and end their videos with longer vowels (even after adjusting for local speech rate). The shortest vowel durations tend to occur in the middle of their videos. However, there are a few notable exceptions, especially in Grishno's videos, some of which demonstrate the exact opposite trend: longer vowels in the middle of a video and shorter vowels at the beginning and end. There does not appear to be a pattern according to video category that governs when or why these exceptions occur.

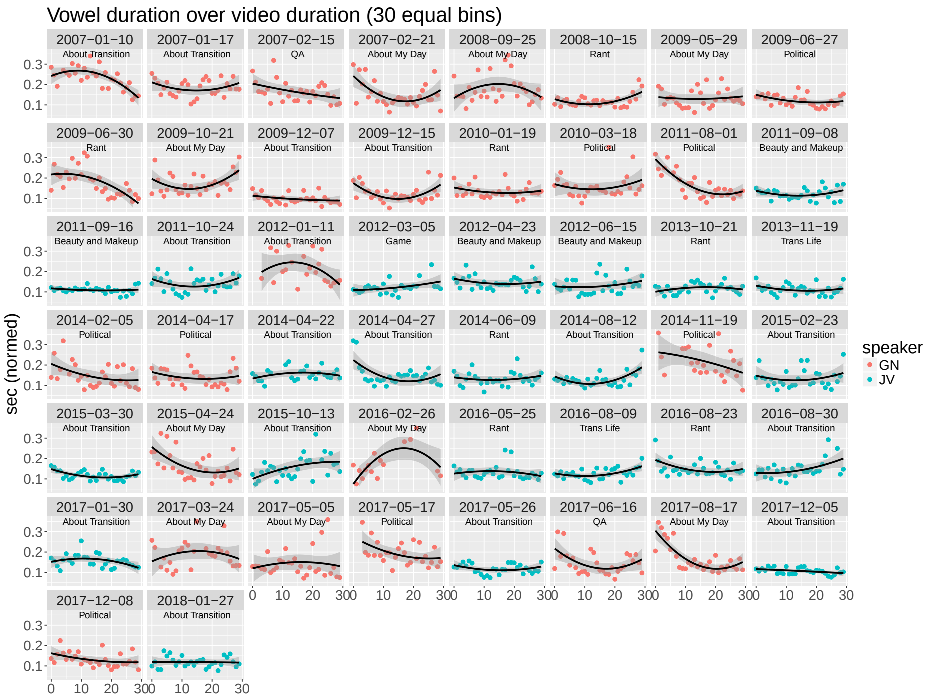

Figure 5: Mean vowel duration, normalized for local speech rate, for each of thirty equal-sized bins of each video, is plotted (color-coded for speaker), along with quadratic regression lines.

5 Discussion

Results from this exploratory analysis show that the two trans female YouTube vloggers demonstrate vastly different patterns of change both chronologically and over the duration of the average video. PrincessJoules' voice has increased in f0 and vowel formant frequencies, which are two of the many sociophonetic variables that have been shown to index a feminine voice. Grishno's voice has decreased in both of these parameters over time.

It would be neither logical nor relevant to say, however, that one speaker or the other is changing to become more or less feminine based on just a few acoustic measurements. It is important to remember that whatever is defined as a feminine voice depends just as much on context as it does on the physical properties of the voice itself. In an in-person discussion with Grishno, she expressed no surprise at the direction in which her voice is trending, and offered a reasonable justification: at the beginning of her transition, before she felt like she could pass as physically female due to her physical appearance, perhaps one of the ways in which she projected her femininity was through a higher-pitched voice and an overall more feminine way of speaking. Indeed, when Grishno spent some time early on in her transition running telephone campaigns, she says that doing dozens of phone calls a day was a perfect time for her to practice different voice styles and registers that would be perceived as female. One decade later and post-transition, Grishno is completely at ease in her body and does not need to depend as much on her voice, hence the decrease in some feminine characteristics.

All this is just a possible explanation for the pattern of one individual, however. PrincessJoules' voice has trended in the opposite direction, and it is safe to assume that no two transgender individuals will really be exactly alike. In her video aptly titled "Feminized Voice"(PrincessJoules, 2014), she explains that she has always had a "higher-pitched... screeching voice", and that this made it easier for her to develop her own voice. She emphasizes "practice, practice, practice" for her viewers who want to achieve a feminine-sounding voice like hers, but also advises that while the voice is an important part of transitioning, it is, of course, subject to change and dependent on context. PrincessJoules' voice has increased significantly in pitch over time, but, in her own words, "who you're with and how comfortable you are with that person also determines how you speak." She gives an illustrative performance with an example of how she might order food at a restaurant in an environment with "lots of guys around and people trying to clock[2]" her.

Figure 6 tracks the fundamental frequency of PrincessJoules' voice just before and during the utterance. This performance can be considered 'hyper-feminine', as she is speaking in a way that is meant to convince or reassure strangers of her feminine identity. The clearest consequence of this is that her pitch rises much higher than her average f0 for the whole video (indicated by the gray dashed line) when she says "Hi, yes...". In addition, however, her speech speed slows considerably, with exaggerated pauses in between each word. In addition to f0 and duration, her enunciation (fully articulated stops and more peripheral vowels) is indicative of a more formal register. This is how PrincessJoules might speak when she consciously tries to index femininity with her voice. With friends and people she's more comfortable

[2] Clocking is the practice of trying to ascertain whether a transgender person is indeed trans by observing them, and is considered rude cisgender behavior in the trans community.

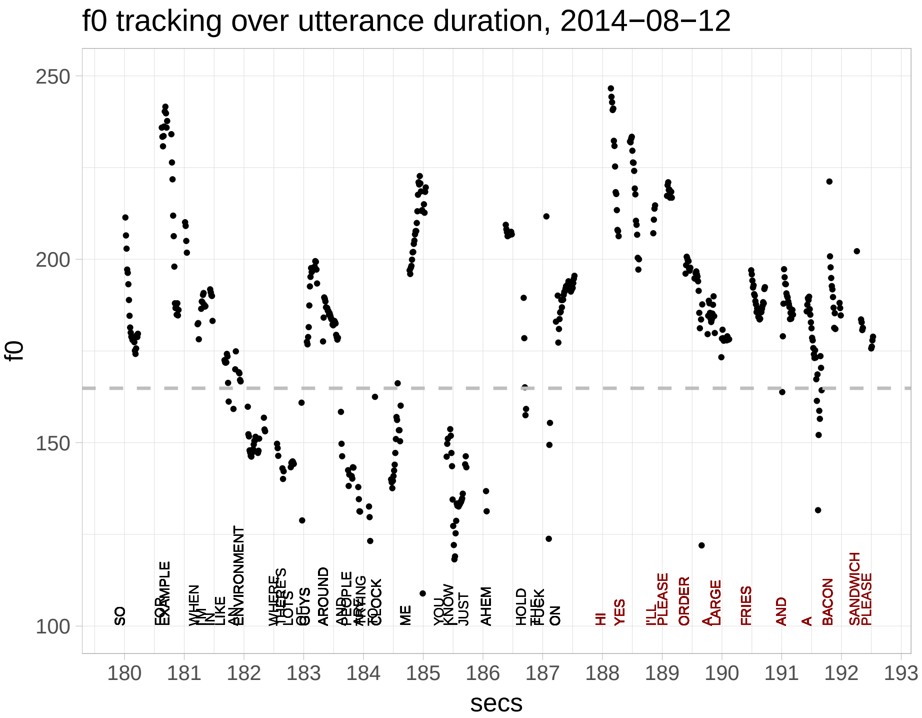

Figure 6: Fundamental frequency tracked over the vowels of each word in an excerpt from PrincessJoules' video "How I Feminized My Voice". The gray dashed line marks her mean f0 for the whole video, and the words in dark red are part of her 'hyper-feminine' performance.

with, on the other hand: "I just let loose, I don't really care, I don't really need to be on guard with my voice."

I highlight these two examples in order to critique the prevailing normative idea that all trans women desire to speak and be perceived as cisgender women of a certain social class and category of femininity in all situations. This is not true (Zimman, 2016; Davis et al., 2014). It is much too simple of a conclusion to make and is not supported by the variation in the data. Trans speakers may have any number of individual, unique goals for their voices during and after their transition, and these goals may change as locally as day-to-day or conversation-to-conversation.

All of this is not to say that those who work with transgender individuals should stop trying to identify specific speech targets for their voices to reach.

I do believe, however, that more research should be done that combines the practice of speech pathology with the theories of individual variation in sociolinguistics. In this corpus of YouTube vloggers in particular, there is much more to be explored regarding gender performativity and style-shifting within an individual video. The way that we advocate for the transgender community through research should include the understanding that all individuals' voices naturally change in different contexts, and that there are many ways to transition that may not all look – or indeed sound – the same.

6 Notes

Thanks to Greer Sullivan, Ronald Sprouse, and Geoff Bacon for their help with transcription and analysis; to the members of UC Berkeley's Phonetics and Phonology Forum for feedback; and of course, endless gratitude to Grishno and PrincessJoules, who do incomparable good work for the LGBTQ community. All errors are my own.

References

Bell, A.
1984. Language style as audience design. *Language in Society*, 13(2):145–204.

Boersma, P. and D. Weenink
2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.40.

Cameron, D. and D. Kulick
2003. *Language and Sexuality.* Cambridge University Press.

Clark, H. H.
1996. Using Language.

Davies, S., V. G. Papp, and C. Antoni
2015. Voice and Communication Change for Gender Nonconforming Individuals: Giving Voice to the Person Inside. *International Journal of Transgenderism*, 16(3).

Davis, J. L., L. Zimman, and J. Raclaw
2014. Opposites attract: Retheorizing binaries in language, gender, and sexuality. In *Queer Excursions: Retheorizing Binaries in language, Gender, and Sexuality*, L. Zimman, J. Davis, and J. Raclaw, eds. Oxford University Press.

Eckert, P. and S. McConnell-Ginnet
2013. *Language and Gender.* Cambridge University Press.

Frobenius, M.
2014. Audience design in monologues: How vloggers involve their viewers. *Journal of Pragmatics*, 72:59–72.

Horak, L.
2014. Trans on YouTube: Intimacy, Visibility, Temporality. *Transgender Studies Quarterly*, 1(4):572–585.

Johnstone, T. and K. R. Scherer
2000. Vocal Communication of Emotion. In *The Handbook of Emotion*, M. Lewis and J. Haviland, eds. Guilford.

PrincessJoules
2014. How I Feminized My Voice [YouTube video].

Raun, T.
2010. Screen-births: Exploring the transformative potential in trans video blogs on YouTube. *Graduate Journal of Social Science, Special Issue:*

Transgender Studies and Theories: Building up the field in a Nordic context, 7(2):113–130.

Raun, T.

2014. Trans as Contested Intelligibility: Interrogating how to Conduct Trans Analysis with Respectful Curiosity. *Lambda Nordica*, 1:13–37.

Rosenfelder, I., J. Fruehwald, K. Evanini, and J. Yuan

2011. FAVE (Forced Alignment and Vowel Extraction) Program Suite [Computer program].

Simpson, A. P.

2009. Phonetic differences between male and female speech. *Language and Linguistics Compass*, 3(2):621–640.

Sprouse, R. L. and K. Johnson

2016. The Berkeley Phonetics Machine. Pp. 1623–1626.

Ueda, Y., T. Hamakawa, T. Sakata, S. Hario, and A. Watanabe

2007. A real-time formant tracker based on the inverse filter control method.

Yuan, J. and M. Liberman

2008. Speaker Identification on the SCOTUS corpus. *Journal of the Acoustical Society of America*, 123(5).

Zabus, C. and D. Coad

2014. Introduction. In *Transgender Experience: Place, Ethnicity, and Visibility*, C. Zabus and D. Coad, eds. Routledge.

Zimman, L.

2016. Sociolinguistic Agency and the Gendered Voice: Metalinguistic Negotiations of Vocal Masculinization among Female-to-Male Transgender Speakers. In *Awareness and Control in Sociolinguistic Research*, A. Babel, ed. Cambridge University Press.