INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Nonbinary American English speakers encode gender in vowel acoustics

Maxwell Hope[1], Charlotte Ward[2], Jason Lilley[2]

[1]University of Delaware, USA
[2]The Nemours Foundation, USA
maxhope@udel.edu, charlotte.ward@nemours.org, jason.lilley@nemours.org

Abstract

Encoding of gender in speech is a well-researched phenomenon, especially as it concerns men and women. Men tend to produce certain acoustic characteristics in certain ways (low fundamental frequency (F0), lower formant frequencies, lower center of gravity for [s] in English) compared to women, though these characteristics also differ based on other group memberships (e.g. race, sexuality, etc). Those who are more feminine, regardless of categorical gender, have been found to produce an increase in F0 and a larger vowel space. However, these previous studies used largely cisgender women and men or only examined encoding of binary gender in speech and did not consider encoding of "other" or nonbinary gender in speech.

This study recruited American English nonbinary speakers to record 400 utterances and correlated acoustic characteristics with multidimensional gender. Masculine, feminine, and "other" gender are significantly correlated with vowel acoustics.

Index Terms: gender expansive, speech production, sociolinguistics, sociophonetics

1. Introduction

1.1. Gender in speech

Previous researchers have attempted to discount differences in speech between people of different genders (e.g. between men and women) as being due to anatomical differences between binary genders or as a result of physiological differences alone. However, they are now better understood as the products of a complex phenomenon that involves social and/or articulatory factors rather than just anatomical factors. Gender cues in speech are influenced by language and culture [1, 2], socialization [3], and individual identity [4], and can change over time [1]. In terms of individual identity, there is a lack of investigation into how those who are gender expansive (e.g. transgender and/or nonbinary) may encode their gender into voice in ways that move beyond a cisheteronormative framework.

1.1.1. Fundamental frequency (F0)

Much of previous research on gender in speech has focused on gender as a binary and has contrasted the differences between the two binary genders: men and women. This research did find some differences between how men and women produce speech, such as in fundamental frequency (F0). Men tend to produce a lower F0 with averages between 107 – 132 Hz, and women tend to produce a higher F0 with averages between 196 – 224 Hz [5].

While previous research on binary genders has brought attention to gender in speech, this research did not account for nonbinary or transgender speech. Recent preliminary research addressing this gap has shown that nonbinary individuals' F0 tends to fall in the middle of men's and women's ranges, with their average being around 144 Hz [6] and with very large variability in F0 production. As this F0 average shows, these voices do not fit into the patterns of male or female voices and their variability defies binary categorization [7].

1.1.2. COG and Peak Frequency of [s] in English

While F0 is indeed a gender-conveying variable for production and perception, it is not the only one. More recent research has found that other variables are also important for conveying gender; in fact, when it comes to gender perception, F0 only accounted for 41.6% of the perceptual ratings of voice gender in one study [8]. Other factors such as center of gravity (COG) of [s] [7, 9] play a role. Similarly to F0, women produce [s] at a higher frequency than men [9-13]. This is measured in either COG or peak amplitude [7]. The typical COG averages for women range between 6,400 – 8,500 Hz, and for men they range between 4,000 – 7,000 Hz [9, 11, 13-17]. In addition to these binary categories, [18] found that in a diverse group of transmasculine speakers, different identities of transmasculinity (e.g. trans men vs nonbinary transmasculine individuals) were encoded in [s] and that this also intersected with queer identities. This group showed what [7] calls a "stylistic bricolage," i.e, the mixing and matching of sociophonetic cues: those who were very masculinely identified but also identified as queer used a low F0 combined with high COG of [s] to signal queer masculinity.

1.1.3. Vocal tract and formant frequencies

Vocal tract length (VTL) and formant frequencies are also known to be correlated with gender. Men tend to produce lower formants in general and women produce higher formants [4, 5]. One thing that influences formants is the VTL; longer vocal tracts result generally in lower formants and smaller vowel spaces, while shorter vocal tracts result in higher formants and larger vowel spaces [4, 19].

Vocal tract manipulation has been used as a technique in speech-language pathology to train transfeminine people to achieve a more feminine voice. One such technique that is used is to encourage the tongue being brought more forward in the mouth (which raises F2) and the lips being spread wider (which shortens the vocal tract, raising all formants, and especially raises F3) [19].

These acoustic variables are not "all or nothing" – they exist on a continuum where individuals can gradiently encode identity. This means that although men may on average have longer vocal tracts than women, both groups can still

 10.21437/Interspeech.2023-1772

manipulate VTL. For example, one study found that the more feminine men are, the larger their vowel spaces were, reflecting a shortening of the acoustic VTL [4]. However, this study did not explicitly include gender expansive speakers and only looked at (presumably) cisgender men and women.

1.2. Questions & hypotheses

Using a binary categorization of gender or a continuum of masculinity to femininity is not conducive to the study of nonbinary voices. Instead a paradigm that contains continuous (scales of masculine, feminine, and other) and categorical (male, female, nonbinary) variables of gender – like the one used in [20], which gave participants both types of variables to describe themselves and the stimuli voices with – is beneficial, as it accommodates the variability of gender expansive voices, provides more opportunities for grouping in analysis, and allows for a more holistic view of the person and their gender.

Our question was broad because of the descriptive nature of this study: how do nonbinary people gradiently encode gender into speech using independent scales of masculine, feminine, and "other" gender? We hypothesized that there would be 1) a positive correlation between feminine gender (i.e. identity and expression) and F0, COG and peak frequency of [s], and vowel formants; and 2) a negative correlation between masculine gender and those acoustic variables. For "other" gender, we anticipated that there would be a more nuanced and complex relationship between other gender and the acoustic variables we investigated.

2. Methods and Materials

2.1. Participants

Gender expansive (e.g. transgender and/or nonbinary) participants over the age of 18 were recruited via word-of-mouth and online to participate in a speech collection study that was approved by the IRB of the first author's affiliation. For this study, we decided to focus only on the subset of participants who identify as nonbinary. Of the sixteen total participants, thirteen nonbinary participants took part. Most of these participants came from the mid-Atlantic region of the United States. They were all fluent speakers of American English who lived in the United States their entire lives. The participants ranged in age from 20 to 34 (mean = 27.1, SD = 4.1). The participants had a variety of genders including but not limited to "transmasculine", "transfeminine", "agender", "genderqueer", and overlapping identities e.g. "nonbinary and transmasculine" or "transmasculine and agender".

2.2. Data collection

Before recording, the participants read and electronically signed an informed consent statement. They then answered a survey with demographic questions relating to gender identity, age, race, and an open-ended question about gender and speech. They provided their "male", "female" and "other" gender *identity* on three independent scales of 0 to 100 and were told that these did not have to add up to 100. They repeated this for "masculine", "feminine", and "other" gender *expression*, resulting in six total gender variables. Next, they proceeded to record ten test sentences in the ModelTalker database [21]. These were used as a screening set to ensure their microphone and their environment were adequate for recording. Once they passed the screening, they recorded the first 400 sentences in the ModelTalker database. Notably, the sentences in the ModelTalker database were chosen and ordered to cover the widest possible range of the most commonly occurring diphones and triphones in English.

2.3. Acoustic analyses

F0 was extracted using a Praat [22] script that took measures at a minimum of 75 Hz and a maximum of 400 Hz for each sentence; then we averaged the F0s for the 400 sentences to get the average F0 per speaker. COG and peak frequency of [s] were extracted using a Praat script; we limited the context to word-initial prevocalic conditions, yielding 39 [s] tokens. The COG and peak frequencies were averaged across all tokens to get each speaker's average COG and peak frequency for [s]. Formants of vowels were extracted using a modified Praat script that takes formant measures at the midpoint of the vowel. In total, each speaker had 371 [i] tokens, 176 [a] tokens, 203 [u] tokens, 201 [e] tokens, and 175 [o] tokens; the formants were averaged per vowel per speaker. Then, using the average formant frequency spacing (Δf) method of [23], we computed the Δf for each speaker and used this value to compute the acoustic vocal tract length (34000/2*Δf).

2.4. Statistical analyses

Statistical analyses were conducted in R [24]. For the correlations and inferential statistics, we converted formant values to their vocal tract normalized values using the method in [23]: Fn' = Fn/ΔF, where *n* is the formant number. This was done to control for differences in anatomy, while preserving sociophonetic differences between speakers.

3. Results

3.1. Descriptive results

3.1.1. Gender Variables

Boxplots showing the distributions of the six gradient gender variables for the group are shown below.

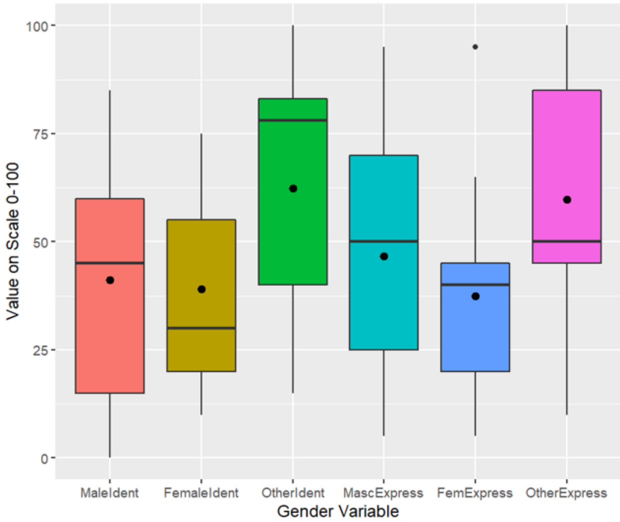

Figure 1: *Boxplots of the six gender variables for the whole group. Bars represent median lines and dots represent means.*

Other gender identity and other gender expression had the highest average values of all six gender variables. For all the values, there was only one participant who rated one variable a 0 and one participant who rated two variables 100; otherwise no other variables were rated at 0 or 100.

3.1.2. F0, COG & Peak Frequency of [s]

The mean and standard deviation (SD) for F0s for the whole group are listed in Table 1 below.

Table 1: *Means and standard deviations for fundamental frequency (F0), COG of [s], Peak Frequency (PF) of [s] in Hz. Reference ranges for binary "male" and "female" speakers are also provided.*

Measure	mean (SD)	Male average	Female average
F0	168.6 (35.5)	107-132 [5]	196-224 [5]
COG	5772.1 (1033.7)	4,757-6,167 [9]	5,727-6,858 [9]
PF	5984.7 (934.3)	4,000-7,100 [11]	6,500-8,100 [11]

The participant with the lowest average F0 (118 Hz) was a nonbinary transmasculine participant who also had the lowest COG and peak frequency for [s] of the group (2584.1 Hz and 3156.7 Hz respectively). The participant with the highest average F0 (224 Hz) was a genderqueer transmasculine participant; however, this person did not have the highest COG or peak frequency for [s]. They ranked fourth from the lowest in terms of COG and peak frequency of [s] (6088.1 Hz and 5775.5 Hz respectively).

3.1.3. Vocal tract acoustics

Average F1, F2, and F3, and computed acoustic VTL for the whole group are shown in Table 2.

Table 2: *Means and standard deviations for the whole group for first three formants in Hz, and the acoustic vocal tract length (VTL) in cm.*

Measure	Mean (SD)
F1	536.1 (33.1)
F2	1767.7 (98.3)
F3	2733.5 (103.2)
VTL	15.1 (0.7)

3.2. Correlations between acoustic variables and gender

Correlations using the Pearson's method were calculated between each of the acoustic variables and each of the gradient gender variables. The acoustic variables consisted of F0, COG, and peak frequency of [s]; overall F1', F2', and F3' values; and average F1', F2', and F3' for each of five vowels ([i], [a], [u], [o], and [e]). F0, COG, and peak frequency of [s] had no significant correlations with any of the gender variables, though the correlation between masculine gender expression and COG for [s] was borderline significant (p = .056, R = -.54). Several of the formant frequency acoustic measures had significant correlations with some gender variables (see Table 3).

3.3. Modeling the data: Linear regression

Significant correlations in Table 3 were modeled with linear regressions using the lme4 package [25] in R. They were visualized using the ggplot function *geom_smooth* using the method *lm* (Figures 2-5). The formula was: *Formant' ~ Gender Variable,* e.g. F2' of [o] ~ Female Gender Identity. For brevity and space, we focus on a few of the F2 results due to the trends and strong correlations.

Table 3: *Significant correlations between normalized vowel formants and gender variables. Values in parentheses indicate Pearson's R; asterisks represent p-value level: p < .01 '**', p < .05 '*'*

Vowel - Formant	Gender Variable
Avg. F2'	Other Identity (.69**)
[i] F2'	Fem. Identity (-.67*)
	Fem. Expression (-.69**)
	Masc. Expression (.61*)
	Other Identity (.69**)
[o] F1'	Fem. Expression (.65*)
[o] F2'	Fem. Identity (.75**)
	Male Identity (-.70**)
	Masc. Expression (-.74**)
[e] F2'	Other Identity (.60*)
	Other Expression (.57*)

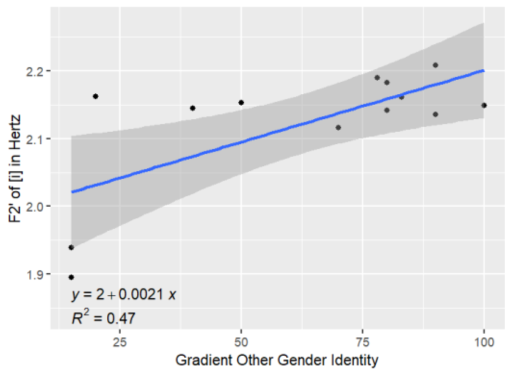

Figure 2: *F2' of [i] in Hertz by Gradient Other Gender Identity.*

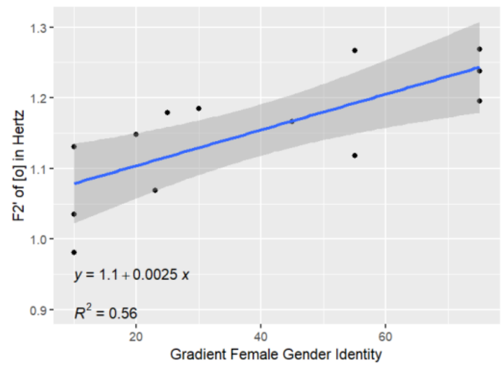

Figure 3: *F2' of [o] in Hertz by Gradient Female Gender Identity.*

4. Discussion

Nonbinary individuals gradiently encode gender in speech. However, contrary to previous research, we did not find significant relationships between feminine gender and F0 or masculine gender and F0. Given the recent findings that gender expansive people did not use F0 as a significant cue to shift the ratings of perceived gender of “ambiguous” voices [20, 26], it makes sense that correlations between F0 and the gradient gender variables would not be significant. We would predict that if we had incorporated binary gender expansive individuals (e.g. trans men and trans women) then we may have found trends with F0, as these individuals may rely on those cues more as a measure to convey gender. The lack of a correlation with COG or peak frequency of [s] was surprising since this is a known sociophonetic cue for gender in English. Although it was not statistically significant at the $p < .05$ level, some of the relationships were borderline significant, with masculine gender expression being the closest to significance: $p = .056$, Pearson’s $R = -.54$. Thus, nonbinary individuals may still be using this cue, in combination with other cues to signal masculine gender expression.

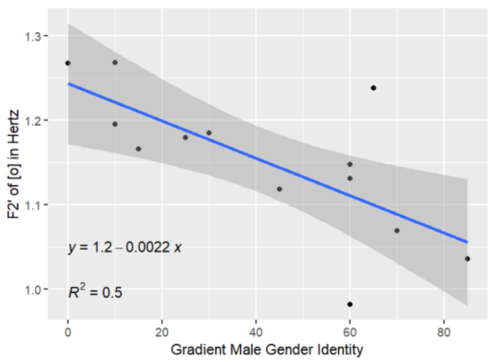

Figure 4: *F2’ of [o] in Hertz by Gradient Male Gender Identity.*

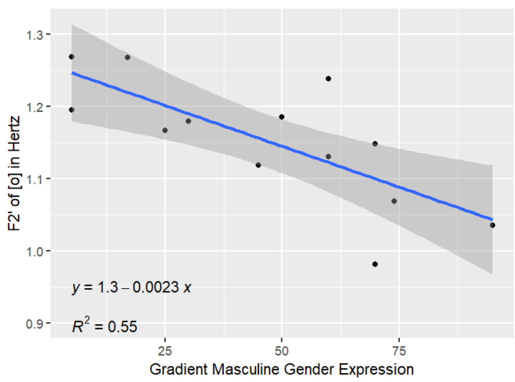

Figure 5: *F2’ of [o] in Hertz by Gradient Masculine Gender Expression.*

After controlling for anatomical differences in the vocal tract using the method of [23], we did find statistically significant correlations between the vowel acoustics and the participants’ gradient genders. This also makes sense given recent findings. In [20], it was found that gender expansive individuals needed vocal tract information to significantly shift their gradient perception of gender (e.g. gradient masculine, feminine, or “other” gender perception) of voices, whereas cisgender listeners only needed the F0 cue. Therefore, gender expansive individuals may be deliberately using manipulation of the vocal tract to subtly and gradiently encode aspects of their gender.

Additionally, voice gender expression may be multidimensional where these speakers are mixing and matching different acoustic properties. For example, it was found that F2 of [o] got significantly lower as gradient male identity increased (Figure 4) and masculine expression increased (Figure 5), with masculine expression having the greater impact of the two. Thus, we have a potential emergent marker of masculinity in this group; however, those who are nonbinary may be gradiently more or less masculine and decide to use other acoustics, such as raising F2 of [i] to signal “other” gender identity (Figure 2) and/or raising other formants to signal simultaneous feminine identity (as these gender variables are not mutually exclusive). Thus, a nonbinary person who has a strong male identity could raise F2 for [i], while lowering F2 of [o]. Another way these speakers may encode simultaneous aspects of gender is by using sibilant production: nonbinary people who embody both masculine and feminine gender may also choose to mix and match a lower F2 of [o] with a higher [s] COG, which is typically perceived as more feminine. In these ways, nonbinary speakers are utilizing “stylistic bricolage” – the mixing and matching of sociophonetic cues – to signal a multidimensional and nonbinary identity.

One limitation of this study was that we were not completely able to control for the environmental factors of the recordings. The participants recorded remotely and although they passed a screening set of test sentences, they all had variable recording environments. Additionally, because we looked only at a small group of nonbinary speakers, this may not be generalizable to a larger group; even for a larger group, nonbinary speakers are diverse both in identities and in speech. Finally, while we modeled the “other” gender variables’ relationships with acoustics using linear regression, this group largely had high “other” gender identity and expression; thus, this analysis was preliminary and further investigation into different ways of modeling this data, that accords with our hypotheses (that “other” gender has a more complex relationship with acoustic variables) is warranted.

5. Conclusion

This study was an exploratory, descriptive investigation into the production of speech of nonbinary individuals and the gradient encoding of multidimensional, nonbinary gender in speech. Thirteen nonbinary speakers recorded 400 English utterances, which were analyzed for multiple acoustic variables known to be correlated with the encoding of gender. We found significant correlations between formants of different vowels ([i], [o] and [e]) and multidimensional gender variables (masculine, feminine, and “other” gender identity and expression). In particular, we uncovered that there was a strong correlation between F2 of [o] and gradient male and female identity as well as gradient masculine gender expression indicating that this group may be using tongue backing of [o] (which drops F2) as a way to signal masculinity. This work has implications for sociophonetics and speech language pathology; these findings could prompt further research into vocal techniques and therapies for gender expansive people of a variety of genders.

6. Acknowledgements

The authors would like to thank Dr. H. Timothy Bunnell for allowing us the resources of the ModelTalker. The authors extend a deep appreciation to every member of the gender expansive community.

7. References

[1] I. P. Yuasa. "Culture and gender of voice pitch: A sociophonetic comparison of the Japanese and Americans." London, UK: Equinox, 2008.

[2] R. van Bezooijen, "Sociocultural Aspects of Pitch Differences between Japanese and Dutch Women," Language and Speech, vol. 38, no. 3, pp. 253–265, 1995, doi: 10.1177/002383099503800303.

[3] C. T. Ferrand and R. L Bloom. "Gender differences in children's intonational patterns," *Journal of Voice,* vol. 10, no. 3, pp. 284–91, 1996, doi: 10.1016/S0892-1997(96)80009-9.

[4] M. Weirich and A. P. Simpson. (2018). "Gender identity is indexed and perceived in speech." *PLoS ONE*, vol. 13, no. 12, 2018, doi: 10.1371/journal. pone.0209226.

[5] S. Davies and J. M. Goldberg. "Clinical aspects of transgender speech feminization and masculinization," *International Journal of Transgenderism*, vol. 9, no. 3–4, pp. 167–196, 2006, doi: 10.1300/J485v09n03_08.

[6] M. Schmid, and E. Bradley. "Vocal pitch and intonation characteristics of those who are gender non-binary," in *Proc. 19th International Congress of Phonetic Sciences*, Melbourne, Australia, 2019, pp. 2685–2689. Available: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_2734.pdf

[7] L. Zimman. "Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/," *Language in Society*, vol. 46, no. 3, pp. 339–370, 2017, doi: 10.1017/S0047404517000070.

[8] Y. Leung, J. Oates, and S. P. Chan. "Voice, articulation, and prosody contribute to listener perceptions of speaker gender: A systematic review and meta-analysis," *Journal of Speech, Language, and Hearing Research*, vol. 61, no. 2, pp. 266–297, 2018, doi: 10.1044/2017_JSLHR-S-17-0067.

[9] S. Fuchs and M. Toda. "Do differences in male versus female /s/ reflect biological or sociophonetic factors?" In Susanne Fuchs, Martine Toda, & Marzena Zygis (eds.), *An interdisciplinary guide to turbulent sounds*, pp. 281–302. Berlin, DE: Mouton de Gruyter, 2010, doi: 10.1515/9783110226584.

[10] M. F. Schwartz. "Identification of speaker sex from isolated voiceless fricatives," *Journal of the Acoustical Society of America,* vol. 43, no. 5, pp. 1178–79, 1968, doi: 10.1121/1.1910954.

[11] P. Flipsen, Jr., L. Shrilberg, G. Weismer, H. Karlsson, and J. McSweeny. "Acoustic characteristics of /s/ in adolescents," *Journal of Speech, Language, and Hearing Research*, vol. 42, no. 3, pp. 663–677, 1999, doi: 10.1044/jslhr.4203.663.

[12] K. Heffernan. "Evidence from HNR that /s/ is a social marker of gender," *Toronto Working Papers in Linguistics,* vol. 23, no. 2, pp. 71–84, 2004. Available: https://twpl.library.utoronto.ca/index.php/twpl/article/view/6208/3197.

[13] J. Stuart-Smith. "Empirical evidence for gendered speech production: /s/ in Glaswegian," in J. Cole and J. I. Hualde Eds., *Laboratory Phonology 9*, pp. 65–86. Berlin, DE & New York, NY, USA: Mouton de Gruyter, 2007.

[14] J. D. Avery, and J. M. Liss. "Acoustic characteristics of less-masculine-sounding male speech," *Journal of the Acoustical Society of America,* vol. 99, no. 6, pp. 3738–3748, 1996, doi: 10.1121/1.414970.

[15] S. Nittrouer. "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," *Journal of the Acoustical Society of America,* vol. 97, no. 1, pp. 520–530, 1995, doi: 10.1121/1.412278.

[16] S. Nittrouer, M. Studdert-Kennedy, and R. S. McGowan. "The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," *Journal of Speech and Hearing Research,* vol. 32, no. 1. pp. 120–1, 1989, doi: 10.1044/jshr.3201.120.

[17] K. Tjaden, and G. S. Turner. "Spectral properties of fricative in amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research,* vol. 40, no. 6, pp. 1358–1372, 1997, doi: 10.1044/jslhr.4006.1358.

[18] L. Zimman. "Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants," *Linguistics*, vol. 55, no. 5, pp. 993-1019, 2017, doi: 10.1515/ling-2017-0018.

[19] L. Carew, G. Dacakis, and J. Oates, "The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals," *Journal of Voice*, vol. 21, no. 5, pp. 591–603, 2007, doi: 10.1016/j.jvoice.2006.05.005.

[20] M. Hope and J. Lilley, "Gender expansive listeners utilize a non-binary, multidimensional conception of gender to inform voice gender perception." Brain and language, 224, 2022, https://doi.org/10.1016/j.bandl.2021.105049

[21] H. T. Bunnell, J. Lilley, and K. McGrath, "The ModelTalker project: A web-based voice banking pipeline for ALS/MND patients," in *Proc. INTERSPEECH 2017 – 18th Annual Conference of the International Speech Communication Association*, Aug. 2017, Stockholm, Sweden, pp. 4032-4033. Available: https://www.isca-speech.org/archive_v0/Interspeech_2017/pdfs/2054.PDF.

[22] P. Boersma and D. Weenink, Praat: doing phonetics by computer [Computer program]. 2022.

[23] K. Johnson, "The ΔF method of vocal tract length normalization for vowels," *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, vol. 11, no. 1, p. 10, Jul. 2020, doi: 10.5334/labphon.196.

[24] R Core Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. 2023. https://www.R-project.org/.

[25] D. Bates, M. Mächler, B. Bolker, and S. Walker. "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software,* vol. 67, no. 1, pp. 1-48, 2015, doi: 10.48550/arXiv.1406.5823.

[26] M. Hope and J. Lilley, "Cues for Perception of Gender in Synthetic Voices and the Role of Identity," in *Proc. INTERSPEECH 2020 – 21st Annual Conference of the International Speech Communication Association*. Oct. 2020, pp. 4143–4147. doi: 10.21437/Interspeech.2020-2657.