Accepted Manuscript
Version of Record at: https://doi.org/10.1016/j.bandl.2021.105049

Gender expansive listeners utilize a non-binary, multidimensional conception of gender to inform voice gender perception

Maxwell Hope[a] and Jason Lilley[b]

[a]University of Delaware, Department of Linguistics & Cognitive Science

125 E Main St

Newark, DE 19716

United States

maxhope@udel.edu

Corresponding author

[b]Nemours Biomedical Research

Center for Pediatric Auditory and Speech Sciences

1701 Rockland Road, Room 136B

Wilmington, DE 19803

United States

jason.lilley@nemours.org

© 2021. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
https://creativecommons.org/licenses/by-nc-nd/4.0/

Accepted Manuscript
Version of Record at: https://doi.org/10.1016/j.bandl.2021.105049

Abstract

Few studies on voice perception have attempted to address the complexity of gender perception of ambiguous voices. The current study investigated how perception of gender varies with the complexity of the listener's own gender conception and identity. We explicitly recruited participants of all genders, including those who are gender expansive (i.e. transgender and/or non-binary), and directed them to rate ambiguous synthetic voices on three independent scales of masculine, feminine, and "other" (and to select one or multiple categorical labels for them). Gender expansive listeners were more likely to use the entire expanse of the rating scales and showed systematic categorization of gender-neutral voices as non-binary. We propose this is due to repeated use of reflective processes that challenge pre-existing gender categories and the incorporation of this decision-making process into their reflexive system. Because voice gender influences speech perception, the perceptual experience of gender expansive listeners may influence perceptual flexibility in speech.

Keywords: Gender perception, perceptual flexibility, gender expansive, transgender, non-binary, voice gender, speech perception

1 Introduction

Gender is a complex social and individual factor that is expressed and processed in a variety of ways, including in the realm of speech perception. The conception of gender itself – such as whether it is binary and fixed or whether it is fluid and multidimensional – also varies from person to person. A multidimensional and non-binary (i.e. not strictly men and women, or masculine and feminine) view of gender is in line with how many gender expansive people conceive of gender (Trans Student Educational Resources, 2015) and reflects the lived experience of the first author of this study, who is gender expansive themself. In this gender framework, "masculinity" and "femininity" exist on separate dimensions of gender, possibly

alongside other dimensions (ibid.). However, studies concerning gender and voice perception have been largely limited to the conceptualization of gender as a binary, by examining stereotypically masculine or feminine voices and recruiting only male or female participants (Skuk & Schweinberger, 2014; Weirich & Simpson, 2018); transgender participants were rarely explicitly recruited. Some studies have expanded their research to include transgender individuals (Junger et al., 2014; Smith et al., 2018), but these studies have primarily used a forced alternative choice paradigm where participants are tasked with saying whether or not the voice sounds “male” or “female”. A recent review by Bent and Holt (2017), who summarized the literature on speech perception with respect to social factors, confirmed that “these task designs are, in some ways, reductionist for many speaker characteristics, because they require listeners to ‘bin’ talkers into specific categories” (p. 2), and invited future speech perception studies to consider gender as more than a binary choice or existing on only one continuum of masculine to feminine. Additionally, these experiments typically involve stimuli where all the vocal parameters are changed or morphed at the same time and to the same degree. For example, fundamental frequency (f_o) and formant frequencies are frequently manipulated in parallel to alter listener perceptions of talker gender (Charest et al., 2013; Junger et al., 2013; Junger et al., 2014), obscuring individual contributions of these parameters to gender perception. Bent and Holt (2017) believed that investigating social factors as multidimensional and continuous not only reflects the populations of interest more accurately but also could solve the ‘lack of invariance’ problem: “social categories can group sounds together that are acoustically distinct as well as separate sounds into distinct categories that are acoustically identical” (p. 2).

1.1 Vocal cues and listener characteristics in gender perception

Different aspects of voice are typically correlated with either a masculine or feminine perception. For example, men and those who have higher levels of testosterone typically have lower f_o due to the thickening of the vocal folds (Dabbs & Mallinger, 1999; Evans et al., 2008; Glaser et al.,

2016). This thickening usually happens during puberty or during the start of testosterone-based hormone treatments, and the corresponding average f_o for this population is 107–132 Hz, with a range of about 80–165 Hz (Davies & Goldberg, 2006). Women and those who have lower levels of testosterone typically have higher fundamental frequencies due to the lack of thickened vocal folds; the average f_o for this population is 196–224 Hz with a range of 145–275 Hz (Davies & Goldberg, 2006). The so-called "androgynous zone", which we refer to as the *neutral range*, has been found to be 140–165 Hz (Davies & Goldberg, 2006). Formant frequencies are also correlated with voice gender, with women/girls typically having higher formant frequencies than men/boys (Davies & Goldberg, 2006; Cartei & Reby, 2013; Leung et al., 2018; Nagels et al., 2020). In a production study which measured the speaker's gender identity on two independent scales of masculine gender and feminine gender, men who rate themselves as less masculine were found to have both higher f_o and larger vowel spaces; in the second part of this study, a voice gender perception experiment showed that there was interaction between the formant frequencies and the listener's gender in determining voice gender (Weirich & Simpson, 2018). Compared to formant frequencies, f_o has been shown to have the stronger impact on voice gender perception. Previous research also found that an "androgynous" f_o contour combined with masculine or feminine formant frequencies significantly influenced voice gender perception (Skuk & Schweinberger, 2014); however, it should be noted that they used a binary choice of perception of "male" or "female". A recent review of the voice gender literature found that pitch only accounted for 41.6% of the variance in perception ratings (Leung et al., 2018). Another factor that influences voice gender perception is intonation. Some studies have shown that women tend to produce more varied pitch with greater falls and rises in pitch compared to men, who tend to produce more flat intonation contours; this does not mean that men are monotone, but that they tend to use primarily flat high and flat low pitch accents rather than rising or falling ones (Clopper & Smiljanic, 2011; Hancock et al., 2014). Finally, characteristics of the listener may influence voice perception. Hancock and Pool (2017) used speech samples from three

different groups (cisgender men, cisgender women, and transgender women) as stimuli in a perception experiment examining whether or not listener characteristics influenced gender and sexual orientation perception; while they found no significant results on perception of the two cisgender groups' speech, they did find that transgender women were perceived as significantly more feminine by "non-straight" cisgender people than by straight cisgender people.

The studies mentioned so far have concerned men and women. While there is little work on non-binary voices, one study (Schmid & Bradley, 2019) showed that speech production differs for non-binary people compared to binary transgender and cisgender individuals. The authors found that non-binary people produce pitch roughly in the middle of men and women (with an average f_o of 144 Hz) and that they also mix f_o contour patterns of both men and women. Thus, so far, there is some evidence in the realm of speech production that binary gender identity does not capture the full picture of speech variability. These differences found in non-binary speech production may impact perception, potentially by creating a new cognitive voice gender category for non-binary speech in the minds of non-binary listeners; however, this has yet to be investigated.

1.2 The current study

Analyzing the previous literature on voice gender production and perception leads us to put the pieces together to see a larger picture: how social factors influence perceptual flexibility. Previous methodological choices may have made it harder to uncover this flexibility with respect to gender. While some studies addressed the limitations of alternative forced choice by using scales (such as Likert or sliding scales) to give femininity or masculinity ratings of voices, these scales similarly rely on a one-dimensional conception of gender (e.g. Wolfe et al., 1990; Hancock et al., 2014). Gallena (2017) added a perception of "gender neutral" categorization, but the listeners were explicitly instructed to use this category for a sample that "could have been

spoken by either a man or a woman" (p. 597), which does not account for or acknowledge the possibility of voices beyond the gender binary or a mix of genders. Additionally, previous studies have not tested gender-ambiguous voices that accurately constrain pitch to the neutral range while other parameters are manipulated. Coleman (1971) used an electrolarynx to control for f_o while manipulating formants; however, f_o was set to a constant 85 +/- 3 Hz which is well below the neutral range (145–160 Hz as identified by Davies & Goldberg, 2006). Gallena (2017) tested speech with pitch held constant at 175 Hz, but this is above previously identified neutral range maxima at 156 Hz (Wolfe et al., 1990), 160 Hz (Davies & Goldberg, 2006), or 165 Hz (Gelfer & Schofield, 2000). Although Gallena (2017) identified 175 Hz as still within the gender-ambiguous range, it is nevertheless 10–19 Hz above previously identified ranges, and therefore could have contributed to their results skewing towards a "not-male" perception.

To address the limitations of previously used scales and categories, we chose to give the listeners in our study more choices, both categorical and continuous (e.g. the option to categorize a voice as multiple genders such as non-binary, genderfluid or agender, and a third scale of "other gender") as a way to probe their possibly multidimensional conception of voice gender. Furthermore, we kept the mean f_o of all stimuli at 144 Hz — closer than Coleman (1971) and Gallena (2017) to the neutral range as identified by previous studies — so that we can more accurately examine the relative contribution of individual vocal tract configurations to gender perception.

Our specific questions of interest are as follows: 1) How do people of various genders perceive gender-ambiguous and less-canonical voices with pitch in the neutral range? 2) For people of various genders, how does their own conception of gender influence their perception of voice gender? 3) Does gradient gender identity which is *between* (e.g. away from 0% or 100%) or *beyond* (e.g. utilization of a third gender scale) the traditional gender binary conception result in voice gender perception which is between or beyond the traditional binary

conception? We sought to answer these questions by implementing a voice gender listening experiment with synthetic voice stimuli. We hypothesized that by keeping pitch in the neutral range, listeners will primarily rely on vocal tract parameters to gender a voice. We further hypothesized a relationship between identity and perception, taking into account categorical and continuous aspects of each, and that this would be influenced by whether or not the individual identifies as gender expansive. Due to their more flexible conception of gender, we hypothesize that those who are gender expansive will be more perceptually flexible than those who are cisgender.

A preliminary analysis of this experiment was presented in Hope and Lilley (2020), using statistical methods that analyzed each of the synthetic voices individually. Here we present a more unified statistical analysis that models all voices at once, and also examine categorical response data that was not previously published.

2 Methods

2.1 Synthetic voice construction

Construction of synthetic voices and stimuli is detailed in Hope and Lilley (2020); we give a summary here. We selected 20 male-identifying and 20 female-identifying speakers from the ModelTalker database (Bunnell et al., 2017) who had each recorded a common corpus of 1589 sentences derived from public-domain literature, and represented a wide range of mean f_o values for both genders. We utilized open-source software (Wu et al., 2009) that can construct synthetic voices from neural-network models. These models take text as input, and output a set of parameters that are fed to a vocoder (Morise et al., 2016) to generate synthetic speech. These parameters can be divided into two sets: those that model the vocal tract, and those that model the f_o contour. We modified the software so that it could train vocal-tract models and f_o-contour models separately. Then, for each type of model, we trained (a) a “male-derived” model

on the corpora of the male-identifying speakers; (b) a “female-derived” model on the corpora of the female-identifying speakers; and (c) a “neutral” model on all 40 speakers. Note that prior to training the male- and female-derived f_o-contour models, we modified the input f_o values by a speaker-specific constant so that their average f_o matched the global mean (144 Hz), but the contour was preserved. Thus, the output synthetic speech would have an overall “neutral” average f_o, but a “male-derived” or “female-derived” contour. We combined the f_o-contour models and vocal tract models to make seven synthetic voices: all possible combinations except the two with “opposite” genders (male-derived combined with female-derived). In addition, we trained two more synthetic voices with unaltered f_o input on (a) the 20 male speakers and (b) the 20 female speakers to make “completely male” and “completely female” voices with mean f_o of about 110 Hz and 175 Hz, respectively (referred to as “MMM” and “FFF” in Hope and Lilley 2020). The nine synthetic voices were used in the perception experiment as outlined below. Note, however, that the “MMM” and “FFF” voices have been excluded from the analysis presented in this paper, since they vary from the other seven voices in mean f_o.

Synthetic voices trained on large amounts of real human speech were used because it takes away the manipulation of the voices by the authors to fit into preconceived parameters. Instead, by letting the model train on female voices, male voices, and a combination of these voices, we were able to create voices that represented “average” versions of a female voice, a male voice, and a “neutral” voice. We checked the outputs of these to ensure they were different in vowel spaces, pitch, and intonation. Additionally, the synthetic voice construction process allowed us to extract separate models of the vocal tracts and intonation contours to mix and match, creating new, ambiguous voices, as described above. The primary advantage of this method is that the models, not the experimenters, decide the parameters based on the training data, removing potential experimenter biases from influencing the stimuli.

2.2 Speech perception survey

As detailed in Hope and Lilley (2020), listeners of all genders, explicitly including those who were gender expansive, were recruited online to participate in an online voice gender perception survey. Listeners first answered various questions about their own gender identity. The first question asked them to indicate their categorical gender from a list of options, where they could select as many options as applied to them; the options were “Man”, “Woman,” “Non-binary”, and “Other” (as a write-in option). Separately, they were also asked whether they considered themselves part of the gender expansive community. Next were three questions that asked the listeners to report their own degrees of feminine, masculine, and “other” *gender identity* on scales from 0 to 100; and three parallel questions about their *gender expression*. Gender identity and gender expression are separate, but often interrelated, aspects of gender as a whole. Gender identity is described as “one’s internal sense of being male, female, neither of these, both, or another gender(s)”, while gender expression is “the physical manifestation of one’s gender identity through clothing, hairstyle, voice, body shape, etc.” (Trans Student Educational Resources, 2015). Because these two aspects often but not always correlate, we felt it important to look at both types of variables as gradients.

After the demographic questionnaire, listeners listened to synthetic stimuli from the nine voices, and rated them on three independent scales (from 0 to 100) of “femininity”, “masculinity”, and “something other than masculinity or femininity.” The listeners rated each voice twice on each scale, with two different sentences as stimuli (see Hope & Lilley 2020 for details). Finally, the listeners were presented with one more sentence for each of the nine voices and asked to check which gender they thought the voice was out of five possible options: Man, Woman, Non-Binary, Agender, and Genderfluid. They were required to check at least one category but were instructed they could check more than one. The sentences used as stimuli were selected from the Harvard Sentences (IEEE, 1969) and are listed in the Appendix.

2.3 Listener demographics

In all, 48 listeners completed the survey. Twenty listeners identified as being part of the gender expansive community; the remaining 28 identified as cisgender. The age range of the whole group was 20 to 69 with an average age of 31.8 (SD = 13.3). One participant who was originally part of the gender expansive group was excluded due to incompletely answering the survey questions and rating all voices as either 0 or 100 on all scales. Thus, 19 remaining gender expansive listeners were included in the final analyses.

2.3.1 Categorical gender

For the cisgender listeners (those who did not identify as part of the gender expansive community), seven identified as men and 20 identified as women. One cisgender listener identified as both a woman and “butch”. None identified as non-binary.

For those who identified as part of the gender expansive community, four identified simultaneously as men and non-binary and one identified as a man as well as genderqueer. There were no men who identified solely as men. Eight gender expansive listeners identified as non-binary, one identified as genderfluid, and one identified as questioning. Finally, four gender expansive listeners identified as women.

2.3.2 Continuous gender

Violin plots showing the distributions of the continuous gender variables for the cisgender and gender expansive groups are provided in the Supplementary Material, along with a summary table. Overall, cisgender participants tended to use the extremes of the scales; the mean scores on the scales matching their gender identity were 75 or above, while the mean scores on the scales of the “opposite” gender were 20 or below, and the means on the “Other” scales were below 8. Also, there were 60 responses of “0” and 17 responses of “100” across all the scales. On the other hand, gender expansive participants tended to use more of the middle range on all

scales, including the “Other” scales; their means were all between 23 and 67. There were only two responses of “0” and none of “100” in this group.

2.4 Statistical analyses of responses to synthetic voices

All descriptive and inferential statistics were computed in R (R Core Team, 2020).

We analyzed the responses using linear mixed effects models, which allow one to model repeated-measures experimental data with multiple independent variables as a multiple regression, while accounting for the non-independence of the data points via the use of the random-effects terms (see e.g. Singmann & Kellen, 2019). To estimate the complex models used here, like the zero-one-inflated models described below, we turned to the R package *brms* (Bürkner, 2017), which uses Bayesian inference. Each model was estimated via Markov chain Monte Carlo (MCMC) sampling, using four chains of 5000 samples apiece; the first 1000 warmup samples per chain were discarded, leaving 16000 samples total for estimation.

We modeled each categorical response option to a voice (Man, Woman, Non-Binary, Agender, and Genderfluid) as an independent binary variable, with a binomial distribution (using the logit link function).

The responses for the continuous response variables (degrees of Masculinity, Femininity, and Other) were bounded to the limits [0,100], with a substantial number of responses at one extreme or the other (particularly 0). Therefore, we divided these responses by 100 and modeled them using a zero-one-inflated beta distribution (Swearingen et al., 2012). Briefly, the beta family of distributions models continuous variables in the bounded range (0,1). A Beta distribution can be fully specified as $B(\mu\phi, (1-\mu)\phi)$, where B is the Beta function, μ is the mean, and ϕ is a “precision” parameter (the higher the value of ϕ, the lower the variance in the distribution). However, the Beta function does not include 0 or 1 in its output, so it cannot model

0 or 1 responses. Hence, we use a zero-one-inflated mixture model where the occurrence of zeroes and ones is modeled separately. In the *brms* package, the model is defined as follows:

$f(x) = \alpha(1-\gamma)$ for $x = 0$

$f(x) = \alpha\gamma$ for $x = 1$

$f(x) = (1-\alpha)\, B(\mu\phi, (1-\mu)\phi)$ for $0 < x < 1$

where α is the zero-one-inflation probability (the probability that zero or one occurs) and γ is the conditional-one-inflation probability (the probability that one occurs instead of zero). In our mixed-effects models, we model all four parameters – μ, ϕ, α, and γ – as dependent on the variables of interest, as elaborated below. The link functions used are log for ϕ (which is positive and unbounded), and logit for μ, α, and γ (which all fall in the range [0,1]).

In what follows, we refer to the listeners' responses about their own gender identities (whether continuous or categorical) as *listener-identity variables* (or *identity variables* for short). Due to the large number of listener-identity variables that could be used as predictor variables for the analyses, the substantial correlations among them, and the relative sparsity of our data, we decided to generate a separate model for each pair of listener-identity variable and response variable. Among the listener-identity variables, self-categorization as a Man, as a Woman, as Non-Binary, as Other, and as gender expansive were each coded as two-level factors, while the six Gender Identity and Gender Expression variables were coded as continuous variables.

Along with a single listener-identity variable, each model also included as categorical predictor variables Vocal Tract Gender (3 levels: Neutral, Male, and Female), f_o-Contour (3 levels: Neutral, Male, and Female), and, for the continuous response variables, Sentence (2 levels for 2 sentences). All factors were coded using treatment contrasts (with the Neutral levels as the references). We found in Hope and Lilley (2020) that Vocal Tract Gender had a far larger

effect on voice gender perception than f_o-Contour. Furthermore, we estimated the amount of variance explained by each of these three variables (calculated with a Bayesian equivalent of R^2 as in Gelman et al., 2019) by comparing a maximal model that included all three variables with models generated with one variable removed. We found that Vocal Tract Gender accounted for about 18-49% of the variance (depending on the response variable), while f_o-Contour and Sentence explained only 3-4% and 5-9%, respectively. Thus, to keep the models relatively simple and computationally tractable, we included an interaction term between Vocal Tract Gender and the listener-identity variable in our models, but did not include interactions for f_o-Contour and Sentence. In the zero-one-inflated beta models, we also modeled α, γ, and ϕ as functions of the identity variable and Vocal Tract Gender, with no interactions. As for random effects, by-listener random intercept and random slope for Vocal Tract Gender was included when modeling the mean and α. Thus, to summarize in the familiar syntax of the R package *lmer*, the complete model for the categorical responses is given by:

Response ~ IdV * VT + f_oC + (VT | Listener)

and for the continuous responses:

μ ~ IdV * VT + f_oC + Snt + (VT | Listener)

α ~ IdV + VT + (VT | Listener)

γ ~ IdV + VT

ϕ ~ IdV + VT

where “VT” indicates Vocal Tract, “f_oC” indicates f_o Contour, “Snt” indicates sentence, and “IdV” indicates the identity variable of interest.

Bayesian models do not produce the equivalent of the *p*-values of frequentist models, but they do produce point estimates of the model parameters, as well as what are known as 95% Credible Intervals, indicating that the model parameter has a 95% probability of lying within the interval. Thus, if the interval does not include zero, we can be reasonably confident that the parameter is non-zero, according to the model. In frequentist terms, such a finding would be called "significant"; below, we use the preferred term "statistically credible".

3 Results

The full results of all models, including coefficients and credible intervals, are provided in the Supplementary Materials. Below, we highlight the main results of interest; unless stated otherwise, all described effects were statistically credible. For clarity, we use SMALLCAPS for all variables, with **RESPONSE VARIABLES** in bold; and the following terminology and labels are used:

- We use the terms Female, Male, and Neutral VOCALTRACT (or VT) to refer to voices built using the female-derived, male-derived and neutral vocal-tract models, respectively; likewise for the Female, Male, and Neutral f_o CONTOUR.

- For the *categorical listener identity variables*, we will refer to the listener's identification as a Man, as a Woman, and as Non-Binary with the labels CATID(Man), CATID(Woman), and CATID(NB) respectively.

- The *continuous listener identity variables* Feminine, Masculine, and Other Gender Identity are labeled as GRADID(F), GRADID(M), and GRADID(O), where "GRAD" is short for "gradient". Similarly, for Gender Expression, we use GRADEXP(F), GRADEXP(M), and GRADEXP(O).

3.1 Categorical perception of voice gender

The *categorical voice response variables* are in **BOLD SMALL CAPS,** using the terminology "voice categorization as a **WOMAN**," for example.

3.1.1 Main effects of VOCALTRACT and f_o CONTOUR on categorical voice gender perception

There was a statistically credible main effect of VOCALTRACT in all models. Unsurprisingly, Male VT voices (compared to other voices) were more likely to be categorized as a **MAN**, and less likely to be categorized as anything else. Similarly, Female VT voices were more likely to be categorized as a **WOMAN**, and less likely to be categorized as a **MAN**, as **AGENDER**, or as **GENDERFLUID**. There was also a credible effect of f_o CONTOUR in some models, with the Female CONTOUR less likely to be categorized as a **MAN**, and more likely to be categorized as a **WOMAN**. Perhaps more surprisingly, in some models the Male CONTOUR was also credibly more likely to be categorized as a **WOMAN** than the Neutral CONTOUR.

3.1.2 Effects of listener-identity variables

For categorical gender perception, there were several statistically credible main effects of identity variables, as well as interactions with VOCALTRACT.

Voice categorization as a **MAN** showed a credibly negative main effect with feminine listener alignment: as either GRADID(F) or GRADEXP(F) variable increased, categorization of the voice as a **MAN** decreased (Figure 1A). Similarly, listeners who identified as Women (CATID(Woman)) were less likely to categorize Male VT voices as a **MAN**, while on the other hand, listeners who identified as Men and/or as Non-Binary were *more* likely to do so.

Accepted Manuscript
Version of Record at: https://doi.org/10.1016/j.bandl.2021.105049

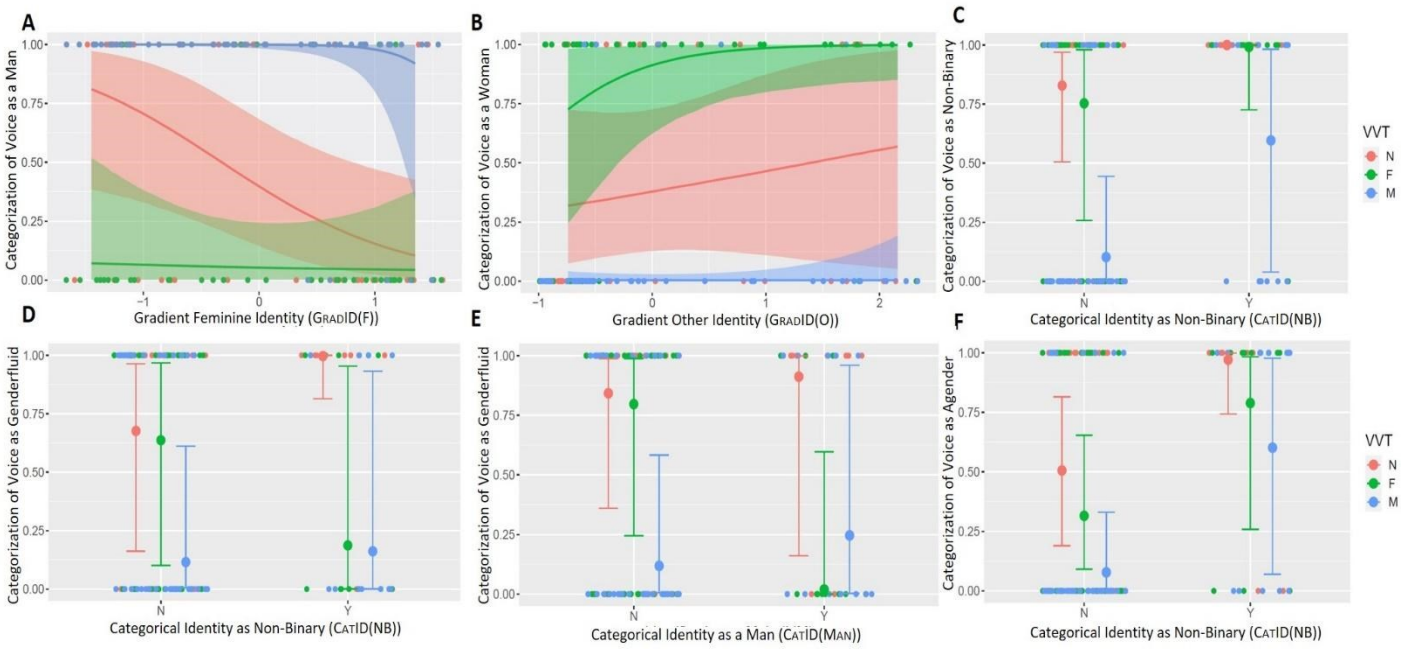


Figure 1. Six plots for categorical gender perceptions: (A) listener's degree of Feminine gender identity vs. voice categorization as a **MAN**; (B) listener's degree of Other identity vs. voice categorization as a **WOMAN**; (C) listener identification as Non-Binary vs. voice categorization as **NON-BINARY**; (D) listener identification as Non-Binary vs. voice categorization as **GENDERFLUID**; (E) listener identification as a Man vs. voice categorization as **GENDERFLUID**; (F) listener identification as Non-Binary vs. voice categorization as **AGENDER**.

There were no main effects associated with voice categorization as a **WOMAN,** but there were three interactions with Female VOCALTRACT. Specifically, Non-Binary listeners (CATID(NB)) and listeners with higher Other Identity (GRADID(O)) were more likely to categorize Female VT voices as a **WOMAN** (Figure 1B), while listeners who identified as Women themselves (CATID(Woman)) were actually *less* likely to do so.

For the other three possible voice categories, there was a notable effect for Non-Binary listeners: they were credibly more likely to categorize a voice as any of **NON-BINARY**,

GENDERFLUID, and **AGENDER** (see Figures 1C, 1D, and 1F). Remarkably, Non-Binary listeners *always* categorized Neutral VT voices as **NON-BINARY** (Figure 1C). However, for the **GENDERFLUID** category, there were credible negative interactions between Female VOCALTRACT and four variables: CATID(NB), GRADID(O), CATID(Man), and GRADEXP(M) – which is to say that Non-Binary listeners (Figure 1D) as well as Men (Figure 1E) were less likely to categorize Female VT voices as **GENDERFLUID**.

3.2 Continuous voice gender perception

In what follows, the *continuous voice response variables* are abbreviated **PERCEPTION(F)**, **PERCEPTION(M)**, and **PERCEPTION(O)**, for Perception of Femininity, Masculinity, and Other, respectively.

3.2.1 Whole-group analysis

3.2.1.1 Main effects of VOCALTRACT and f_o CONTOUR on continuous voice gender perception

As with the categorical models, there was a statistically credible main effect of VOCALTRACT in all continuous perception models. As expected, the Female VT voices had the highest **PERCEPTION(F)** ratings and lowest **PERCEPTION(M)** ratings, while the opposite was true for Male VT voices. For **PERCEPTION(O)**, the Neutral VOCALTRACT was rated highest, and credibly different than the Male VOCALTRACT, which was lowest (the Female VT, in between, was not credibly different from either). The Female f_o CONTOUR was credibly rated higher for **PERCEPTION(F)** and lower for **PERCEPTION(M)**, but there was no effect for the Male CONTOUR; and there was no effect of CONTOUR on **PERCEPTION(O)**.

3.2.1.2 Effects of identity variables on continuous voice gender perception

Surprisingly, we found no statistically credible main effects of any identity variables on **PERCEPTION(F)**, nor interactions of those variables with VOCALTRACT. However, there were several credible main effects on **PERCEPTION(M)**. In particular, there were positive effects of both GRADEXP(F) and GRADID(F), as well as negative effects of both GRADEXP(M) and GRADID(M) (Figure 2A); in short, listeners who were more *masculinely* aligned perceived the voices as *less* masculine than those who were more *femininely* aligned – although for GRADID(M), this effect did not hold for Female VT voices (cf. Figure 2A). On the other hand, an interaction between CATID(Man) and VOCALTRACT (Figure 2B) showed that Men credibly rated the Neutral VOCALTRACT as less masculine than the Male VOCALTRACT.

For **PERCEPTION(O)**, there was a statistically credible main effect of both GRADID(O) (Figure 2C) and CATID(NB) (Figure 2D), showing that Non-Binary listeners and those with higher "Other" gender identities rated the voices as credibly more “Other” than other listeners did, which falls in line with the results for Non-Binary speakers with categorical perception in Section 3.1.2.

We also note that these two variables, CATID(NB) and GRADID(O), as well as GRADEXP(O) and identification as gender expansive, each had credible negative effects on the α parameter, the probability of a 0 or 100 score, on either two or all three of the **PERCEPTION** scales. These effects all indicate that these (overlapping) listener groups (Non-Binary, gender expansive, and Other) were less likely to rate a voice as either a 0 or 100 overall. In fact, there were only 10 scores of “100” in the gender expansive group (which included all Non-Binary

listeners) versus 38 in the cisgender group, and only 57 scores of “0” versus 241 in the cisgender group.

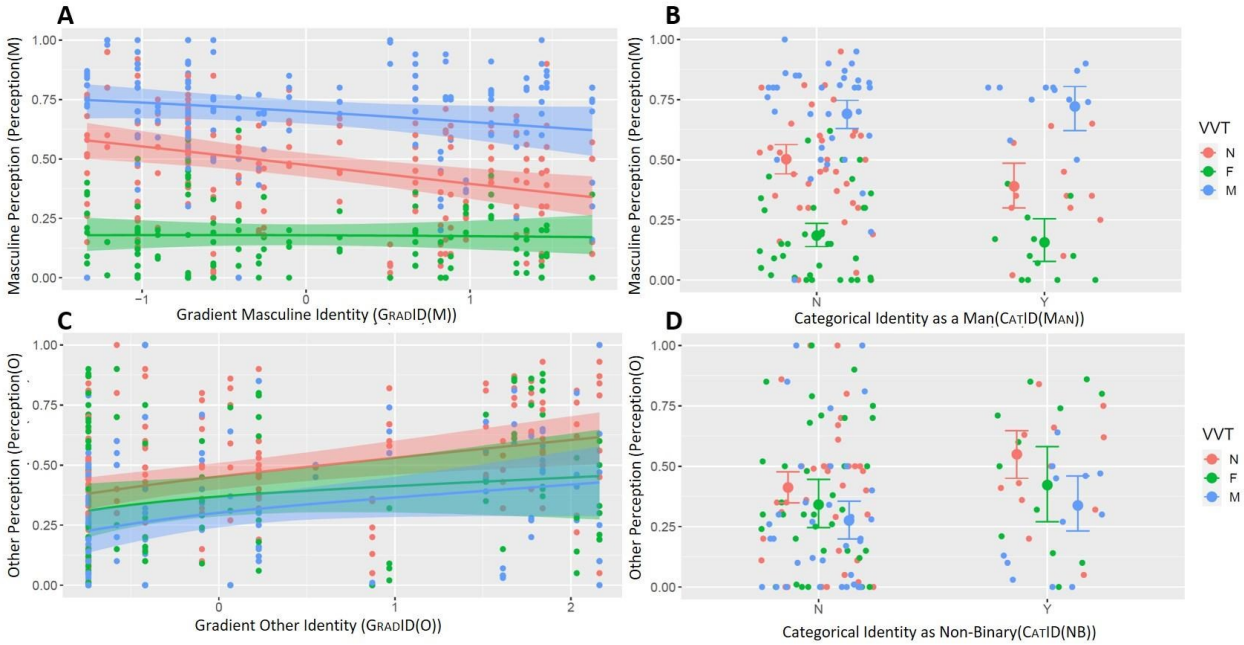


Figure 2. Four mixed effects model plots for continuous gender perceptions: (A) masculine gender identity vs. masculine gender perception; (B) categorical listener identification as a Man vs. masculine gender perception; (C) other gender identity vs. other gender perception; (D) categorical listener identification as Non-Binary vs. other gender perception.

3.2.2 Group analysis: Cisgender vs gender expansive listeners for continuous voice gender perception

In addition to the all-listeners models discussed above, models were also estimated for the Cisgender and gender expansive groups separately for the **Perception** scales. These models were identical in structure to the whole-group models, except that γ was removed from the models for the gender expansive group, due to the lack of scores of “100” in this group.

3.2.2.1 Main effects of VocalTract and f_o Contour on continuous voice gender perception by group

The effects of VocalTract and f_o Contour were largely the same for the two subgroups as for the whole group (see 3.2.1.1), but fewer of the measured differences were statistically credible. There was no credible VocalTract effect for **Perception(O)** in some Cisgender models, while in some gender expansive models, there were credible effects of both Male and Female VT voices on **Perception(O)**. The effect of the Female f_o Contour on **Perception(F)**, seen in the whole group and Cisgender models, was not credible in the gender expansive models.

3.2.2.2 Effects of identity variables on continuous voice gender perception by group

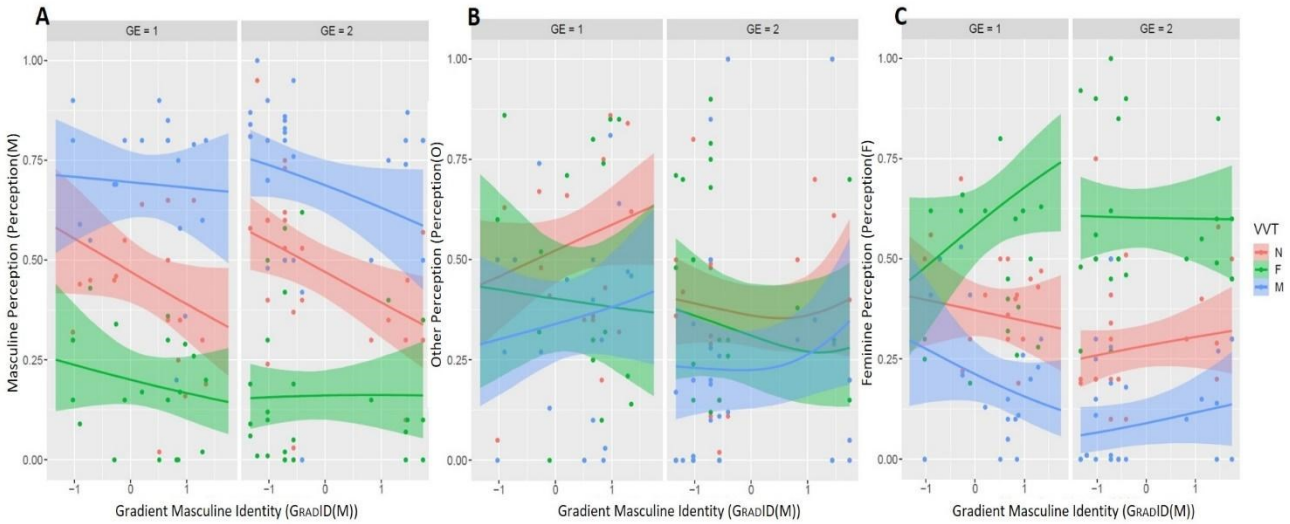
There were several differences between the two groups in the effects of the identity variables—in fact, we found no credible effects or interactions in common between them.

The Cisgender listener group showed the same main effects on **Perception(M)** as described above for the whole group, namely that **Perception(M)** was positively correlated with listener femininity (GradExp(F) and GradID(F)) and negatively correlated with listener masculinity (GradExp(M) and GradID(M)) (see Figure 3A, right). Also, unique to the Cisgender group was a negative correlation between **Perception(O)** and Masculine Gender Identity (GradID(M)) (Figure 3B, right).

For the gender expansive group, as was found in the whole group, there was a credible interaction between CatID(Man) and VocalTract on **Perception(M)**: the Male VocalTract was rated as credibly higher in **Perception(M)** than the Neutral VT by Men. Also, though no effects were found on **Perception(F)** in the whole group or Cisgender group, the gender expansive group had one: as Masculine Identity (GradID(M)) increased, the **Perception(F)** of the Female VocalTract increased (Figure 3C, left). Finally, we point out again the results

found in Section 3.2.1.2 for Non-Binary listeners (who were all gender expansive); namely, that they used higher scores on the **PERCEPTION(O)** scale, and were less likely to give a "0" or "100" score.

Figure 3. Mixed-effects model plots for continuous perception, showing gender expansive listeners on the left and cisgender on the right: (A) masculine identity vs. masculine perception;



(B) masculine identity vs. other perception; (C) masculine identity vs. feminine perception.

4 Discussion

This study sought to explore several questions concerning the differences between non-binary, multidimensional gender conception and binary, one-dimensional gender conception and how these differences impact voice gender perception. We addressed 1) how people of various genders perceive gender-ambiguous and less-canonical voices with pitch in the neutral range, 2) how people of various genders incorporate their own conception of gender to influence their perception of voice gender, and 3) if gradient gender identity which is *between* or *beyond* the traditional gender binary conception results in voice gender perception which is between or

beyond the traditional binary conception. The next three subsections will discuss continuous and categorical perceptions by listener identity and alignment.

4.1 Men and masculinely-aligned listeners' perceptions

Among men and those more masculinely-aligned in general, a notable contrast was found. For *continuous* perception of masculine gender, listeners with a higher degree of masculine identity rated voices as *less* masculine than ones with a lower degree of masculinity – although this was statistically credible only in the cisgender group. This may indicate that cisgender men are more conservative in their perception of masculinity. On the other hand, when it came to *categorical* perception, the Male VOCALTRACT was *more* likely to be categorized as a Man by those who categorically identified as men; and this was not correlated with degree of masculine identity. Our interpretation of this finding is that it helps to highlight the difference between categorical gender and gradient gender; it is the difference between "which box you put yourself in" and "how well the box fits". Belonging or not belonging to the group is more relevant to the categorization of the voice as a man, possibly indicating strong in-group versus out-group effects. Listener's categorical identity as a Man also had a strong negative relationship with categorical perception of the voice as Genderfluid of the Female VOCALTRACT voices; this perhaps indicates that for Men, their idea of a Genderfluid voice is mutually exclusive with a Female VOCALTRACT, making the Genderfluid voice category more restrictive than Agender or Non-Binary voice gender categories.

We also note that those who were gender expansive and had a higher degree of masculine gender identity showed *higher* ratings of femininity for the Female VOCALTRACT voices. This effect was not found for cisgender men or cisgender masculinely-aligned individuals, and in fact was notable for being the only credible effect on perception of femininity. This finding may indicate that gender expansive masculinely-aligned people are more sensitive

to femininity in voice, which could be due to underlying barriers that non-binary and masculine gender expansive people face in being appropriately gendered; they have to have a clear idea of what sounds feminine so that they can distance themselves from it when and if they wish to.

4.2 Women and femininely-aligned listeners' perceptions

Women and femininely-aligned listeners seemed to show opposite but parallel trends to the ones discussed above for men and masculinely-aligned listeners, regarding perception of masculinity and categorization as a man. On the one hand, those who rated themselves higher in femininity tended to rate voices as *more* masculine – although, again, this trend was only credible in the cisgender group. On the other hand, the Male VOCALTRACT was *less* likely to be categorized as a Man by those who categorically identified as women. But note that women were also less likely to categorize the Female VOCALTRACT voices as Women. Taking these results together, one might say that women are more generous in their ratings of masculinity than men, but are also more likely to utilize other voice gender categories (e.g. Non-Binary, Agender, or Genderfluid) *instead of* the Man and Woman categories.

A plausible conclusion is that women are more flexible in categorizing voices outside of their traditional voice gender percepts. It may also be that women have, in general, stronger perceptual boundaries for what a man's or woman's voice sounds like.

4.3 Non-binary and other-aligned listeners' perceptions

Several results indicated that listeners who were non-binary and those who had higher other gender identity were more likely to use the middle range of the masculine and feminine perception scales and also more likely to use labels outside the traditional binary; for example, they rated voices higher on the Other perception scale. The Neutral VOCALTRACT in particular was categorized as Non-Binary 100% of the time, and was also categorized credibly more often

as Genderfluid or Agender. On the other hand, these listeners were also more likely than other listeners to categorize Male VOCALTRACT voices as a Man and Female VOCALTRACT voices as a Woman. Considering all these facts together, it can be speculated that those with greater Other identity and those who identify as non-binary show a third category of "other" gender voices, distinct from male and female voices. This category maps onto a neutral vocal-tract anchor – that is, a reference point for what non-binary voices sound like. It may be that, because they have a salient third gender-category, non-binary listeners and those who are other-aligned in gender will show different categorical perceptions of speech sounds.

4.4 Neurocognitive framework of voice gender perception & proposed model

One of the predominant models that attempts to account for the perception of social factors is a dynamic model, set forth by Michael Lieberman and others (Lieberman et al., 2002; Lieberman et al., 2004; Satpute & Lieberman, 2006; Lieberman, 2007), which integrates the reflexive (X) and reflective (C) systems in the brain. According to this theory, perceivers make use of the X-system which automatically assesses the similarity of the other person to oneself when making judgements about others; it works to find a match between the input and a label and primarily recruits the lateral temporal cortex. If it cannot adequately label a target based on pattern recognition, the C-system is activated primarily in the prefrontal cortex (Lieberman, 2007). This helps the perceiver to discover a new solution, which in turn is fed back to the X-system so that when it encounters similar targets in the future, it may be able to come to a conclusion before the C-system has to be activated again. In short, the X-system makes use of automatic pattern recognition and where it fails, the C-system helps to find new solutions. These new solutions can either reinforce the previous patterns of the X-system or help to store new patterns. Specific to voice gender perception, Charest et al. (2013) proposed a two-stage model based on fMRI findings. Relating their findings to the X- and C-system approach, they found that there was increased temporal cortex activity for voice stimuli towards the ends of the masculine-to-

feminine spectrum, indicating possible X-system use, and increased brain activity in the prefrontal cortex when presented with ambiguous stimuli – which could point to utilization of the C-system – since "ambiguous voices were more difficult to rate as male or female, less categorically defined as one or the other gender, thus requiring more energy for decision making" (ibid., p. 965).

We propose that acoustic cues are first automatically processed by the temporal cortex and then mapped onto existing patterns in the prefrontal cortex. When a target voice gender is more congruent with the perceiver's gender, there is a facilitatory effect causing the existing pattern in the prefrontal cortex to be more activated while other patterns are suppressed. The relationship between the target voice gender and the perceiver's gender varies based on the social dynamics at play; voice gender perception therefore relies on the individual's conception of gender as well as the societal "in-group" and "out-group" dynamics between people of various genders. A more complex and non-exclusive conception of gender could lead to an increase in speech perceptual flexibility, which is known to be influenced by the listener's gender (Bent & Holt, 2017).

The relationships discovered between identity and perception can be interpreted in the context of perceptual flexibility in speech. When people shift their phonetic boundaries, they can incorporate new sounds, such as non-native sounds, into their perception. The moving of the boundaries can either encompass more phonetic variation of a phone (Sjerps & McQueen, 2010) or it can make room for a new category, as can happen with second language learning (Francis & Nusbaum, 2002). Similarly, our results show that those listeners who are cisgender and more feminine have broadened their category of what masculine voices sound like and included the neutral voices in their perception of masculinity. Those who reported greater "other" identity are more similar to language learners who create a new category for second language acquisition, except that in this case, they have an additional voice gender category. Both of

these may contribute to flexibility in speech perception. This is also supported by previous findings that showed that auditory encoding can be shaped from the top-down processing of higher-level knowledge and attention (Heald & Nusbaum, 2014), and processing the complexity of gender and challenging previously held beliefs about gender is precisely the type of higher-level process that could impact processing of acoustic cues. This invites us to interweave the X/C-system theory into speech perception to better understand perceptual flexibility in speech.

Figure 4 presents a proposed neurocognitive model of voice gender perception based on the integration of our results with the neurocognitive models discussed. We propose that the C-system, the reflective system of the brain which utilizes more controlled processing, carries the biases about what different gendered voices sound like in the prefrontal cortex, specifically the medial and dorsomedial prefrontal cortices. When presented with speech, the acoustic cues are processed in the auditory ventral stream and pattern-matched by the X-system, the reflexive system responsible for automatic processing, in the lateral temporal cortex. If a match fails, such as in the presentation of an ambiguously gendered voice, the anterior cingulate cortex is triggered to find a new solution. Either this solution can utilize only the extremes of a gradient variable, thereby reinforcing the previous conceptions held in the prefrontal cortex, or it can create a new category in the middle range of the variable, thus utilizing the variable's full continuum. In the former case, the decision process influences the C-system so that it becomes increasingly easier to pattern-match that voice in the future; in the latter case, there will be an increase in the perceptual flexibility of the speech sound, which also influences the C-system over time.

Accepted Manuscript
Version of Record at: https://doi.org/10.1016/j.bandl.2021.105049

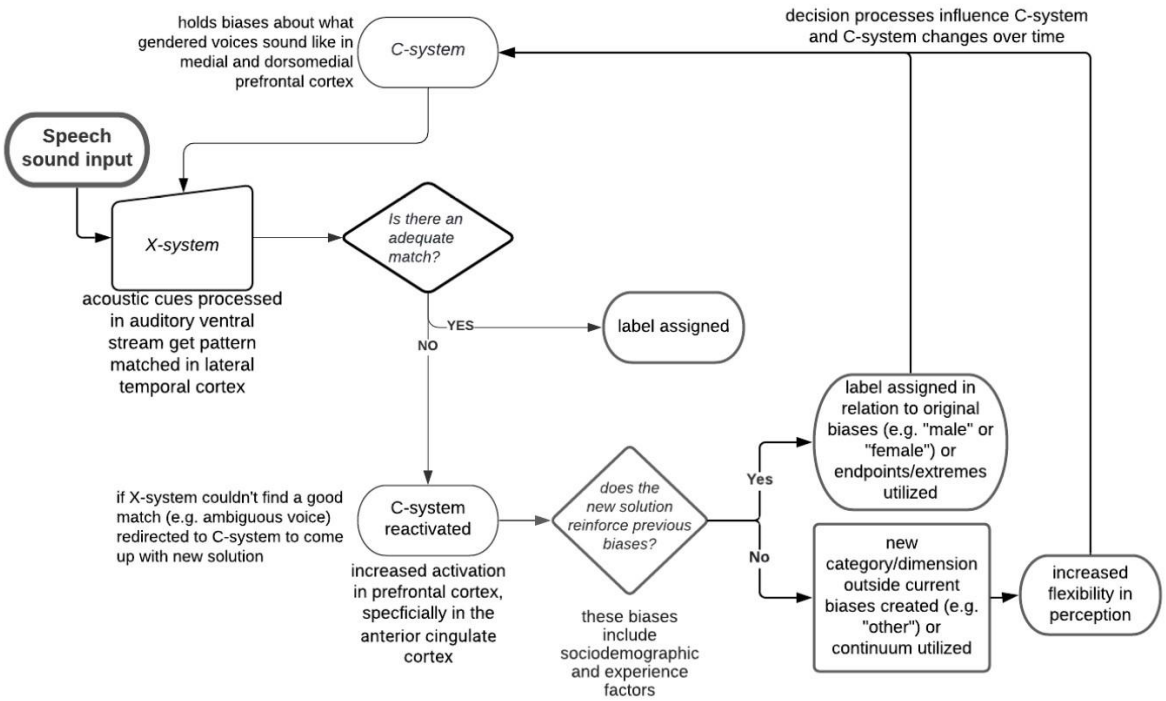


Figure 4. Proposed neurocognitive model of voice gender perception with neural correlates, adapted from Lieberman et al. (2002), Lieberman (2007) and Charest et al. (2013)

Our interpretations of our results based on this model are that gender expansive listeners have recruited more of the C-system in the past where they have made decisions to label voices in new ways that have challenged their preconceptions. After this repeated decision-making process over time, these decisions have been integrated into their C-system, which has modified the potential pattern matches they are capable of within the X-system. They have incorporated a broader range of voices into specific categories, so as to facilitate more automatic processing of less-canonical voices. We hypothesize that in future neurocognitive studies of voice gender, this facilitation effect will be especially notable in those who identify as non-binary specifically since they demonstrated a separate category for neutral voices, using the neutral vocal tract range as an anchor for a non-binary category of voice gender. Cis women have likely also integrated broadened voice gender categories into the C-system but have likely

not added a distinct third category for neutral voices. Therefore, we hypothesize this would make their perceptions of masculine and feminine gender of voices, including ones with neutral vocal tracts, more automatic compared to cis men, though likely not as automatic as gender expansive people. Finally, cis men show the most rigid categories when it comes to feminine gender and masculine gender. With our proposed model, we speculate that cis men, on average, have not challenged their preconceptions and categories of gender, labeling voices as "male" or "female" and resulting in continued use of the C-system when there are less-canonical voices, and therefore less reflexive processing of less-canonical voices. We predict from this that in future neurocognitive studies, cis men will show the slowest response times and increased brain activation to ambiguously gendered voices, as has been shown in previous studies.

We can now reinterpret various results in the literature in light of our findings and model. A previous neurocognitive study on voice gender perception by Junger et al. (2014) argued that trans women have trained themselves to have a more feminine voice in order to explain their findings that trans women had no greater increase in brain activity to feminine voices than to masculine voices. Our interpretation of their results is that trans women are more perceptually flexible due to acknowledging, and adapting to, the fact that people have a wide range of vocal characteristics, and that this is not inherently linked to any one gender. This is also the more trans affirming interpretation, as not all trans women conform to stereotypical voice gender production. A study by Smith et al. (2018) found that trans men are more accurate than cis men at processing masculine components of ambiguous male voices, and had decreased processing load when doing so. The authors state that this may be due to them being sensitive to the "aspired sex"; however, within the context of our findings, it is more likely that trans men have expanded their category of what masculinity can sound like due to their own conception of gender.

One of the limiting factors in this study was that we did not assess the listeners' degree of familiarity with the gender expansive community. Some gender expansive listeners may be new to the community, while some of the cisgender listeners may have close friends or family who are gender expansive. This level of familiarity could impact the results we found. One other aspect that could be investigated in terms of familiarity is to put a "name" to a "voice" – that is, to personalize the voices by, for example, presenting them with written descriptions of people as they listen to the voice (e.g. "This is Quinn. Quinn is a university student and has three siblings"). This could be insightful, as a previous study on face perception showed that "personalizing experiences, even minimal experiences, inhibit group perception processes for outgroup members" (Zárate et al., 2008, p. 113); thus personalization could also prove to increase flexibility of gender perception in voice, leading listeners to categorize voices under multiple possible genders. Another limitation of the study is the artificialness of the stimuli. We did not intend for the participants to believe they were listening to real human voices (in fact, we informed them in advance that the stimuli were synthetic), so no naturalness ratings of the stimuli were collected. It is possible that people respond differently to noticeably artificial stimuli, and in the future, it could be worth investigating this again with more "authentic" sounding stimuli.

5 Conclusion

This paper has explored one social factor, gender, as a multi-dimensional variable that is used to process perception of voice gender. It has also demonstrated that social factors such as gender are important to perceptual flexibility overall. These findings are consistent with work in other domains such as in race which shows that those who are biracial are overall more cognitively flexible with respect to race categorization than monoracial individuals (Pauker & Ambady, 2009). Our findings are also consistent with research about impression formation, which show that the characteristics of the *perceiver* contribute more to forming impressions than

the appearance or presentation of the *target* (Xie et al., 2019), and that a person's conceptual knowledge of self is reflected in their social judgements of others (Stolier et al., 2020).

Researchers can utilize this more complex understanding of gender and this voice gender perception framework in future studies to examine neural correlates of voice gender perception for various populations, including the gender expansive population. Doing so will pave the way for better understanding of the impacts of social factors on perceptual flexibility.

Acknowledgements

We would like to thank Dr. Kathryn Franich and the two anonymous reviewers for their extensive commentary and advice on earlier drafts of this paper. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Bent, T., & Holt, R. F. (2017). Representation of speech variability. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(4), 1–14. https://doi.org/10.1002/wcs.1434

Bunnell, H. T., Lilley, J., & McGrath, K. (2017). The ModelTalker project: A web-based voice banking pipeline for ALS/MND patients. In *Proceedings of the 18th Annual Conference of the International Speech Communication Association (INTERSPEECH 2017),* 4032–4033.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cartei, V., & Reby, D. (2013). Effect of formant frequency spacing on perceived gender in pre-pubertal children's voices. *PLoS ONE*, *8*(12), 12–18. https://doi.org/10.1371/journal.pone.0081022

Charest, I., Pernet, C., Latinus, M., Crabbe, F., & Belin, P. (2013). Cerebral processing of voice gender studied using a continuous carryover fMRI design. *Cerebral Cortex*, *23*(4), 958–966. https://doi.org/10.1093/cercor/bhs090

Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics, 39*(2), 237–245. https://doi.org/10.1016/j.wocn.2011.02.006

Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research, 14*(3), 565–577. https://doi.org/10.1044/jshr.1403.565

Dabbs, J. M., & Mallinger, A. (1999). High testosterone levels predict low voice pitch among men. *Personality and Individual Differences, 27*(4), 801–804. https://doi.org/10.1016/s0191-8869(98)00272-4

Davies, S., & Goldberg, J. M. (2006). Clinical aspects of transgender speech feminization and masculinization. *International Journal of Transgenderism*, *9*(3–4), 167–196. https://doi.org/10.1300/J485v09n03_08

Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior, 93*(4–5), 783–788. https://doi.org/10.1016/j.physbeh.2007.11.033

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349–366. https://doi.org/10.1037/0096-1523.28.2.349

Gallena, S. J. K., Stickels, B., & Stickels, E. (2017). Gender perception after raising vowel fundamental and formant frequencies: Considerations for oral resonance research. *Journal of Voice, 32*(5), 592–601. https://doi.org/10.1016/j.jvoice.2017.06.023

Gelfer, M. P., & Schofield, K. J. (2000). Comparison of acoustic and perceptual measures of voice in male-to-female transsexuals perceived as female versus those perceived as male. *Journal of Voice, 14*(1), 22–33. https://doi.org/10.1016/S0892-1997(00)80092-2

Gelman, A., Goodrich, B., Gabry, J. & Vehtari, A. (2019). R-squared for Bayesian regression models. *The American Statistician, 73*(3), 307–309. https://doi.org/10.1080/00031305.2018.1549100

Glaser, R., York, A., & Dimitrakakis, C. (2016). Effect of testosterone therapy on the female voice. *Climacteric, 19*(2), 198–203. https://doi.org/10.3109/13697137.2015.1136925

Hancock, A. B., Colton, L., & Douglas, F. (2014). Intonation and gender perception: Applications for transgender speakers. *Journal of Voice, 28*, 203–209. https://doi.org/10.1016/j.jvoice.2013.08.009

Hancock, A. B., & Pool, S. F. (2017). Influence of listener characteristics on perceptions of sex and gender. *Journal of Language and Social Psychology*, *36*(5), 599–610. https://doi.org/10.1177/0261927x17704460

Heald, S. L., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience, 8*. https://doi.org/10.3389/fnsys.2014.00035

Hope, M., & Lilley, J. (2020). Cues for perception of gender in synthetic voices and the role of identity. In *Proceedings of the 21st Annual Conference of the International Speech Communication Association (INTERSPEECH 2020)*, 4143–4147.

IEEE. (1969). Harvard Sentences. *Subcommittee on Subjective Measurements: IEEE Recommended Practices for Speech Quality Measurements. IEEE Transactions on Audio and Electroacoustics, 17*, 227–46.

Junger, J., Habel, U., Bröhr, S., Neulen, J., Neuschaefer-Rube, C., Birkholz, P., . . . Pauly, K. (2014). More than just two sexes: The neural correlates of voice gender perception in gender dysphoria. *PLoS ONE, 9*(11). https://doi.org/10.1371/journal.pone.0111672

Junger, J., Pauly, K., Bröhr, S., Birkholz, P., Neuschaefer-Rube, C., Kohler, C., Schneider, F., Derntl, B., & Habel, U. (2013). Sex matters: Neural correlates of voice gender perception. *NeuroImage*, *79*, 275–287. https://doi.org/10.1016/j.neuroimage.2013.04.105

Leung, Y., Oates, J., & Chan, S. P. (2018). Voice, articulation, and prosody contribute to listener perceptions of speaker gender: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, *61*(2), 266–297. https://doi.org/10.1044/2017_JSLHR-S-17-0067

Lieberman, M. D. (2007). The X- and C-Systems: The neural basis of automatic and controlled social cognition. In E. Harmon-Jones & P. Winkielman (Editors), *Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior* (pp. 290–315). New York, NY: Guilford.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, *34*, 199–249. https://doi.org/10.1016/s0065-2601(02)80006-5

Lieberman, M. D., Jarcho, J. M., & Satpute, A. B. (2004). Evidence-based and intuition-based self-knowledge: An fMRI study. *Journal of Personality and Social Psychology*, *87*(4), 421–435. https://doi.org/10.1037/0022-3514.87.4.421

Morise, M., Yokomori, F., & Ozawa, K. (2016). WORLD: a vocoder-based high-quality speech synthesis system for real-time applications. *IEICE Transactions on Information and Systems, E99-D* (7), 1877–1884.

Nagels, L., Gaudrain, E., Vickers, D., Hendriks, P., & Başkent, D. (2020). Development of voice perception is dissociated across gender cues in school-age children. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-61732-6

Pauker, K., & Ambady, N. (2009). Multiracial faces: How categorization affects memory at the boundaries of race. *Journal of Social Issues, 65*(1), 69–86. https://doi.org/10.1111/j.1540-4560.2008.01588.x

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*(1), 86–97. https://doi.org/10.1016/j.brainres.2006.01.005

Schmid, M., & Bradley, E. (2019). Vocal pitch and intonation characteristics of those who are gender non-binary. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 2685–2689). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. In D. H. Spieler & E. Schumacher (Eds.), *New Methods in Cognitive Psychology* (pp. 4–31). Psychology Press.

Sjerps, M. J. & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 195–211. https://doi.org/10.1037/a0016803

Skuk, V. G. & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language, and Hearing Research*, *57*(1), 285–296. https://doi.org/10.1044/1092-4388(2013/12-0314)

Smith, E., Junger, J., Pauly, K., Kellermann, T., Neulen, J., Neuschaefer-Rube, C., . . . Habel, U. (2018). Gender incongruence and the brain – behavioral and neural correlates of voice gender perception in transgender people. *Hormones and Behavior, 105*, 11–21. https://doi.org/10.1016/j.yhbeh.2018.07.001

Stolier, R. M., Hehman, E., & Freeman, J. B. (2020). Trait knowledge forms a common structure across social cognition. *Nature Human Behaviour*, *4*(4), 361–371. https://doi.org/10.1038/s41562-019-0800-6

Swearingen, C., Melguizo Castro, M., & Bursac, Z. (2012). I*nflated beta regression: Zero, one, and everything in between.* Paper presented at SAS Global Forum 2012: Statistics and Data Analysis, paper 325.

Trans Student Educational Resources. (2015). “The Gender Unicorn.” www.transstudent.org/gender

Weirich, M., & Simpson, A. P. (2018). Gender identity is indexed and perceived in speech. *PLoS ONE, 13*(12). https://doi.org/10.1371/journal.pone.0209226

Wolfe, V. I., Ratusnik, D. L., Smith, F. H., & Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *Journal of Speech and Hearing Disorders, 55*(1), 43–50. https://doi.org/10.1044/jshd.5501.43

Wu, Z., Watts, O., & King, S. (2009). Merlin: An open source neural network speech synthesis system. *Proceedings of the 9th ISCA Speech Synthesis Workshop (SSW9)*, 218–233. http://ssw9.talp.cat/download/ssw9_proceedings.pdf

Xie, S. Y., Flake, J. K., & Hehman, E. (2019). Perceiver and target characteristics contribute to impression formation differently across race and gender. *Journal of Personality and Social Psychology*, *117*(2), 364–385. https://doi.org/10.1037/pspi0000160

Zárate, M. A., Stoever, C. J., MacLin, M. K., & Arms-Chavez, C. J. (2008). Neurocognitive underpinnings of face perception: Further evidence of distinct person and group perception processes. *Journal of Personality and Social Psychology*, *94*(1), 108–115. https://doi.org/10.1037/0022-3514.94.1.108

Accepted Manuscript
Version of Record at: https://doi.org/10.1016/j.bandl.2021.105049

Appendix

Harvard Sentences (IEEE, 1969) used for stimuli:

1. The birch canoe slid on the smooth planks.
2. Glue the sheet to the dark blue background.
3. It's easy to tell the depth of a well.
4. These days a chicken leg is a rare dish.
5. Rice is often served in round bowls.
6. The juice of lemons makes fine punch.
7. The box was thrown beside the parked truck.