

Random bisexual forests: Intersections between gender, sexuality, and race in /s/ production

Chloe Willis & Chadi Ben Youssef*

Abstract. The ideology of “the gay lisp” has inspired numerous quantitative studies examining the relationship between /s/ production and sexuality in American English (e.g., Linville 1998; Munson et al. 2006a; Zimman 2013). There are two key gaps in this literature. First, research in this area typically focuses on monosexual (i.e. lesbian, gay, and straight) speakers to the exclusion of bisexuality. Second, work in this area rarely considers the intersection of sexuality with factors outside of gender or sex, and to a lesser extent geographic location (e.g., Campbell-Kibler 2011; Podesva & Hofwegen 2014). This article addresses these disparities by (1) centralizing bisexual speakers and (2) attending to social factors such as race, place, age, and their intersections in the analysis. To that end, we build upon previous work by Willis (forthcoming) and apply a random forest (Breiman 2001) to /s/ center of gravity measurements. In doing so, we follow Tagliamonte & Baayen (2012) in demonstrating the utility of random forests as an approach to quantitative sociolinguistic analysis. Ultimately, the analysis underscores the need to attend to power structures and biases within research practice: the monosexist ideologies of sexuality and gender normativity that obfuscate bisexuality, and the privileging of whiteness that permeates quantitative studies of sexuality and the voice.

Keywords. bisexuality; random forests; research practice; /s/-fronting; center of gravity; sexuality and the voice

1. Introduction. There is a widespread belief in the United States that English-speaking gay men “talk like women” (e.g. speak with higher pitch, pronounce /s/ in a way that sounds “lispy”) and that lesbian women “talk like men” (e.g. speak with lower pitch and/or monotone intonation). Language and sexuality researchers have invested considerable effort in attempting to prove or disprove whether these stereotypes have any “real” basis in speech production. There are two key gaps in this literature. First, research in this area typically focuses on monosexual (i.e. lesbian, gay, and straight) speakers to the exclusion of bisexuality. However, Willis (forthcoming) demonstrates that bisexual speakers do not pattern consistently with lesbian/gay or straight speakers with respect to /s/ production. Second, work in this area rarely considers the intersection of sexuality with factors outside of gender or sex, and to a lesser extent geographic location (e.g., Campbell-Kibler 2011; Podesva & Hofwegen 2014). However, Calder and King's (2020) recent work with Black communities in California and New York demonstrates that /s/ variation is linked to both speaker race and geographic location. The current study addresses these gaps in the literature through a quantitative analysis that (1) centralizes bisexual speakers and (2) attends to social factors such as race, place, age, and their intersections.

To that end, we iterate on previous work by analyzing the data presented in Willis (forthcoming) using a non-parametric machine-learning technique, namely random forests (Breiman 2001). Willis presents a linear mixed-effects regression model of /s/ center of gravity to analyze the production tendencies of bisexual English speakers relative to lesbian, gay, and straight speakers.

* Authors: Chloe Willis, University of California Santa Barbara (chloemwillis@ucsb.edu) & Chadi Ben Youssef, University of California Santa Barbara (chadi@ucsb.edu).

The presented analysis contains valuable insights and adheres to methodological standards established in similar work in the area. However, the scope of the analysis is constrained by a number of factors. Specifically, the models include only a limited number of predictors due to issues of sample size and structure. Indeed, these problems are quite common among datasets used in linguistics, which are typically assembled through observation or experiments. These data collection methods tend to engender certain characteristics, such as:

1. non-random sampling
2. small sample size
3. repeated-measures structure
4. an imbalanced dependent variable
5. too many predictors relative to sample size
6. low cell counts or empty cells

It is often the case that linguistic datasets are compiled using non-random sampling methods, such as convenience or snowball sampling. This lack of randomness potentially leads to a non-normal distribution of the data. In such cases, standard parametric tests that assume a normal distribution, such as logistic and linear regressions, are not appropriate for analysis because the normality assumption is violated from the onset. Furthermore, a small sample size may also lead to cases in which there are few data points relative to the number of predictors and interactions that may be of interest. Even less extreme versions of these so-called “small N large P ” cases can lead to cell counts that are too low for reliable parameter estimation when using standard parametric models (Strobl et al. 2009). Regardless of sampling method or sample size, linguistic datasets usually involve repeated measures. That is, participants contribute more than one data-point in most linguistic studies. Statistical tests that assume observations are independent (e.g. analysis of variance or ANOVA) are not suitable for analyzing data with a repeated-measures structure. For sociolinguists who rely on observational or experimental data, it is crucial to adopt a modeling technique that overcomes the inherently imbalanced nature of such datasets for hypothesis generation, testing, and modeling. We argue that random forests are better equipped for this task compared to more “traditional” modeling techniques.

A random forest (Breiman 2001) is a tree-based supervised machine learning algorithm used to identify structure in the relations between a response variable and multiple predictors for regression and classification tasks, among others. Random forests contain a predetermined number of trees. For each tree in the forest, the algorithm repeatedly divides the data into successively smaller groups based on the values of the dependent variable. Each division or split within a tree leads to the largest improvement possible in terms of classification or prediction accuracy for the response variable (Gries 2013; Ben Youssef & Gries forthcoming). Once all of the individual trees are grown, a random forest reports a final prediction that is the mean or mode of the predictions from all of the trees (Figure 1).

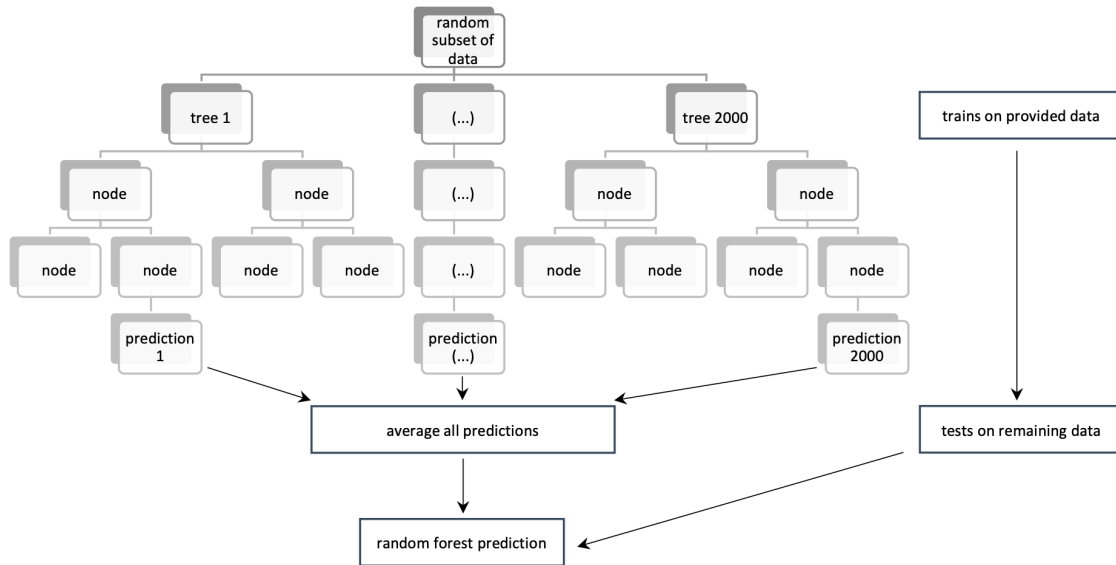


Figure 1. Random forest example

What makes random forests suitable for data with the above characteristics (1-6)? Random forests are a type of nonparametric test and therefore do not assume a particular data distribution, normal or otherwise. Therefore they are appropriate for cases in which the data are not normally distributed or the sample size is too small to make an educated guess about the true shape of the distribution (1, 2). Moreover, random forests are relatively resistant to the effects of empty or low cell counts, making them suitable for “small N large P ” cases and/or datasets that are unbalanced due to non-random sampling (1, 2, 5, 6). Next, random forests do not assume independence of the observations and thus are effective at dealing with repeated measures (3). Finally, random forests are often superior when predicting an imbalanced response variable, i.e. one that is characterized by an uneven distribution of its levels (4) (Muchlinski et al. 2016).

In addition, random forests add two layers of randomness that produce a number of advantages compared to more “traditional” modeling methods in linguistics. These layers of randomness are not included by default in other methods, which often require rigorous post-hoc validation. The first layer of randomness happens at the tree level (e.g. tree 1 in Figure 1). Individual trees are grown using *bootstrapping* or random sampling with replacement. In other words, only a random subset of the data is considered when growing a tree (as opposed to the entire dataset). The data set aside at this stage is later used to test the accuracy of or to *validate* the results of the tree (e.g. “tests on remaining data” in Figure 1). Thus, unlike other modeling approaches commonly used in linguistics, such as linear regression or ANOVA, model validation is built into the process of growing a random forest. The second layer of randomness occurs at the level of the split (the transition from tree to node in Figure 1). For each split in a tree, a random forest considers a randomly selected subset of predictors. Together with bootstrapping, limiting splits “decorrelates trees, helps identify the importance of predictors and their interactions to the predictions, avoids collinearity problems, and protects against overfitting” (Ben Youssef & Gries forthcoming).

ing:9).¹² To summarize, the data and predictors each tree “sees” are randomized, which makes separating the wheat from the chaff and generalizing beyond a given dataset more straightforward and less prone to misleading interpretations.

All of this is not to say that other modeling techniques are wholly inappropriate for (quantitative socio-) linguistic questions. Rather, random forests are a viable option when the idiosyncrasies of the data make other modeling approaches problematic or limiting. In this article, we demonstrate the application of random forests in the context of quantitative sociolinguistic research, namely, the relationship between social factors and /s/ production.

2. /s/. Variation in /s/ production has been linked to a number of identity categories and social characteristics in North-American, English-speaking contexts. From gender, to sexuality, region, and class, the /s/ sound has a complex web of interrelated social or indexical meanings (e.g., Munson 2007; Campbell-Kibler 2011; Podesva & Hofwegen 2014; Calder 2021) (Figure 2). This voiceless anterior fricative is produced by creating a narrow constriction between the tongue and the alveolar ridge, such that the airstream passes over the back of the top teeth. Differences in /s/ articulation are often discussed in terms of a cline of frontness [+] to backness [-]. On the front end of the spectrum, the tongue is placed relatively close to the teeth (i.e. in the front of the mouth), which creates a high frequency hissing sound. On the back end of the spectrum, the tongue is retracted away from the teeth (i.e. towards the back in the mouth), creating a lower frequency hissing sound (Calder 2021; Fuchs & Toda 2010).

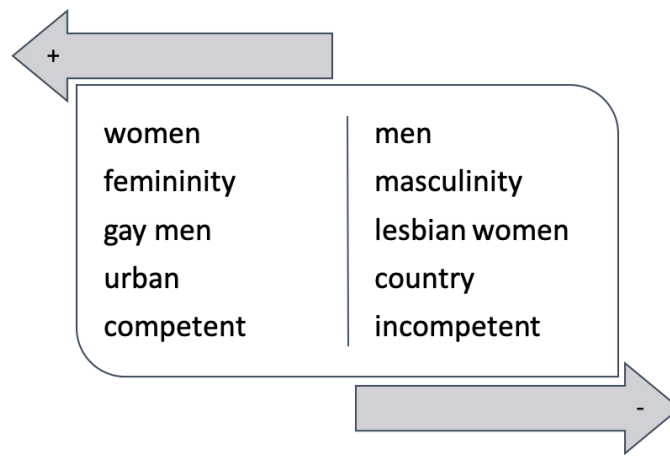


Figure 2. Indexical fields of fronted [s+] and retracted [s-] (adapted from Calder (2021))

These differences in tongue position are connected to an acoustic correlate referred to as center of gravity in the literature. Center of gravity (COG), refers to the weighted mean of frequencies in the spectrum.³ It typically ranges from 6.4 kHz to 8.5 kHz for English-speaking women and from 4 kHz to 7 kHz for English-speaking men (Zimman 2017). Studies of English speakers suggest that gay men produce /s/ with higher COG than straight men (e.g., Linville 1998;

¹ We take “collinearity problems” to refer to, for example, multicollinearity or a phenomenon in which one predictor variable in a multivariate regression model is linearly predicted by another predictor variable at above-chance levels. Multicollinearity is problematic when determining what variables are important for the analysis.

² Overfitting refers to a situation in which an analysis corresponds too closely to a particular data set, which is problematic when the goal is to make predictions that apply beyond the training data.

³ Center of gravity is sometimes also known as the centroid or as the (average) peak frequency.

Munson et al. 2006a; Podesva & Hofwegen 2014). Perceptually, higher COG /s/ tokens are consistently evaluated as more gay-sounding and less masculine-sounding relative to low COG tokens (Linville 1998; Rogers & Smyth 2003; Munson & Babel 2007; Campbell-Kibler 2011; Zimman 2013). Studies of women’s speech are less coherent, with some research finding that lesbian women produce lower COG /s/ than straight women (Munson et al. 2006a; Hazenberg 2015) and others finding no significant difference (Barron-Lutzcross 2015). Regardless, there appears to be relatively reliable distinctions in /s/ vis--vis gender and sexuality, whether they be described in articulatory (e.g. front vs. back) or acoustic (e.g. high vs. low COG) terms.

The relationship between /s/ center of gravity, gender, and sexuality is further inflected by axes of social difference such as competency, regional affiliation, age, and race. Campbell-Kibler (2011) reports a complex web of indexical associations between /s/ frontedness (as measured by COG), masculinity, sexual orientation and competency. Specifically, she reports that listeners tend to cluster higher COG /s/ with “intelligent, effeminate, gay men” on the one hand and lower COG /s/ with “unintelligent, masculine, straight men” on the other. She also finds a relationship between /s/ fronting and regional variation, such that lower COG /s/ tokens are associated with southern U.S. dialects and sounding “country”. Likewise, Podesva and Hofwegen’s (2014) study of English speakers in Redding, California reports that /s/ production varies relative to gender, sexuality, age, and orientation to “country” (i.e. a socially conservative political ideology defined in opposition to “city” or urban life). Gay men in Redding produce higher COG /s/ than straight men in their community, albeit within the limits of local gender norms. /s/ fronting also inversely correlates with age in this population, such that older speakers produce lower COG /s/, and this effect is driven by country-oriented participants. Finally, work by Calder and King (2020) discusses the relationship between /s/, gender, and region in Black communities. They report gender-based distinctions in /s/ frontedness among African American speakers in urban Rochester, New York, but do not find such differences among African American speakers in non-urban Bakersfield, California. In short, /s/ variation as measured by COG has a complex indexical field (Eckert 2008) of interrelated social meanings related to gender, sexuality, region, age, and race.

We continue in the tradition of sociolinguistic research on English /s/ and fit a random forest to a dataset of /s/ center of gravity measurements. The methods used to collect the data analyzed here (that is, lab-based experiments with participants unfamiliar to the researchers) limit our ability to further theorize the indexical field of /s/. Regardless, the previous literature on the social meanings of /s/ influenced our decision to examine these particular measures, and thereby laid the groundwork for this study. Rather than further discussing the indexical meanings of /s/ variation, we instead focus on the implications of our findings for research practice in our discussion.

3. Data.

3.1. **SPEAKERS.** Twenty-seven cisgender native English speakers from varying ethnoracial backgrounds and places of origin were recruited from a California university (Table 1). Speakers ranged from 18 to 30 years old at the time of recording and self-identified as bisexual ($n = 7$), lesbian ($n = 5$), gay ($n = 5$), or straight ($n = 10$). Participants were recruited using snowball sampling and through flyers distributed electronically and posted in various physical locations on the university campus.

Group	n	Mean age	Age range	Race/ethnicity	Place of origin
Bisexual women	5	21	18-26	1 Latinx/PI, 4 white	US
Bisexual men	2	19	18-19	1 Latinx, 1 white	US
Lesbian women	5	21	19-26	1 Latinx/white, 3 white, 1 white/Jewish	3 US, 2 US & Europe
Gay men	5	25	18-30	1 Asian, 1 Latinx/white, 3 white	4 US, 1 India
Straight women	5	19	18-20	2 Asian, 1 Latinx, 2 white	4 US, 1 declined
Straight men	5	19	18-20	3 Latinx 2 white	US

Table 1. Participant information

3.2. **STIMULI.** Participants were recorded reading the Rainbow Passage (Fairbanks 1960), a phonetically balanced scientific passage about rainbows, in a private, sound-attenuated booth using Audacity (Audacity Team 2019). They were instructed to read the passage twice only and as naturally as possible. Next, /s/ tokens were extracted from participants’ first readings of the passage using a Python script (Zimman 2018). The script generated measurements for the dependent variable.

4. Design. We fit a random forest on the collected data using the *randomForestSRC* package (Ishwaran & Kogalur 2022) and the *ggRandomForests* package (Ehrlinger 2016) in *R* (R Core Team 2022). The predictors were as follows: GENDER, RACE, FEMININITY RATING, MASCULINITY RATING, SEXUALITY, REGION, AGE, WORD, and SPEAKER. Self-reported gender stereotypicality ratings were coded as ordinal variables with 7 levels (1 = not at all feminine/masculine, 7 = very feminine/masculine). RACE was coded such that mixed identities were conflated into a single level (e.g. Latinx/white). Participants’ places of origin were also conflated into regions for interpretability: East coast, West coast, Southwest, Midwest, Rocky Mountains, India, multiple, and declined. SPEAKER and WORD function similarly to random effects in mixed-effects models in that they theoretically capture any individual differences between participants or structural differences between words that are not covered by the other named predictors in the forest.

We fit the COG random forest using hyperparameters that performed optimally during the development stage: the number of trees to grow in the forest, *ntree*, and the number of randomly sampled predictors to use in each tree, *mtry*. Table 2 summarizes the model’s hyperparameters as well as the resulting error rates and R^2 (the proportion of variation in the dependent variable that is predictable from the independent variables) for the COG model. To contextualize model error rate, we also include the error rate of a baseline model fit as a single tree with only the most important predictor, GENDER (for the notion of importance, see below).

Hyperparameters		Model statistics		
<i>n</i> tree	<i>m</i> try	Baseline error rate (COG ~ GENDER)	RF error rate	R^2
900	6	1382926	693492	0.56

Table 2. Hyperparameters and model statistics for COG random forest

When reporting and interpreting the results of a random forest, it is crucial to examine variable importance scores as well as partial dependence plots (Gries 2013). Variable importance scores (VIMP) reflect the absolute size of the effect of a predictor on the response variable. In a random forest used for a regression task (such as the one reported here), VIMP scores represent the equivalent of how far regression coefficients of (z-standardized) predictors are from zero in whatever direction. Figure 2 (below) plots VIMP scores computed by randomly permuting each variable’s values and comparing the prediction error to that of the observed values. A large VIMP score suggests that the variable is important for obtaining accurate predictions, whereas a VIMP score closer to zero suggests that the variable contributes little to the accuracy of predictions.

We employ a joint-variable approach (Ishwaran 2007) to detect potential interactions between predictors. In this approach, the paired importance of each pair of variables is calculated and then subtracted from the sum of both variables’ VIMP scores. A large association value indicates that the interaction is worth exploring (but not necessarily significant) when the univariate VIMP score for both variables is also relatively large. What constitutes a “high” or “low” association value is decided by the researcher, similar to the process of setting a significance threshold (Ben Youssef & Gries forthcoming). Here we report interactions with relatively large association values that are of theoretical importance to this analysis.

5. Results. Figure 3 features the VIMP scores for COG. GENDER, RACE, REGION, MASCULINITY RATING, and FEMININITY RATING all have a relatively large effect on the forest’s predictions and make substantial contributions to prediction accuracy. The low scores for SPEAKER, SEXUALITY, WORD, and AGE indicate that there is little variation in COG across these factors in this sample.

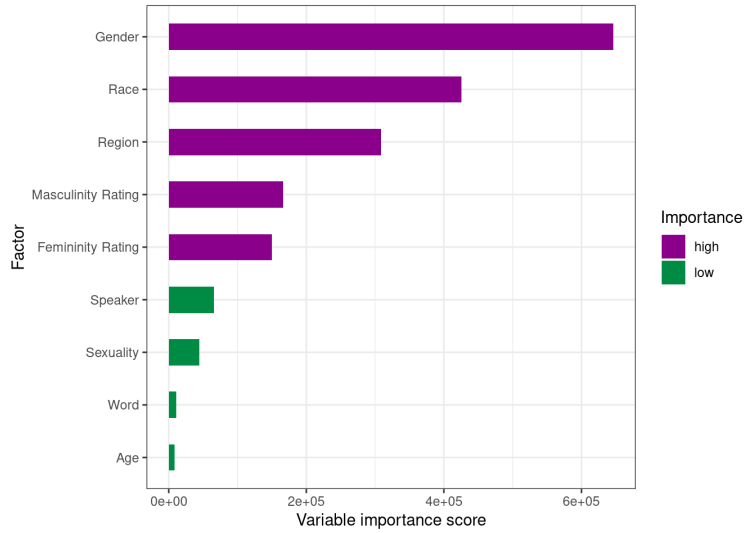


Figure 3. COG VIMP scores

As for interactions, findings indicate that GENDER:RACE is the most important interaction when predicting COG in this sample (Figure 4). Note that not all GENDER:RACE comparisons are represented, as no Latinx/PI or white/Jewish men participated in the study. Women produced higher COG estimates than men across all ethnorracial categories in the sample. The gap between predicted COG is largest among monoracial Latinx women and men at approximately 2.5 kHz. Gender-based differences in predicted COG are more moderate between monoracial Asian and white speakers and rather minimal among mixed race Latinx/white speakers.

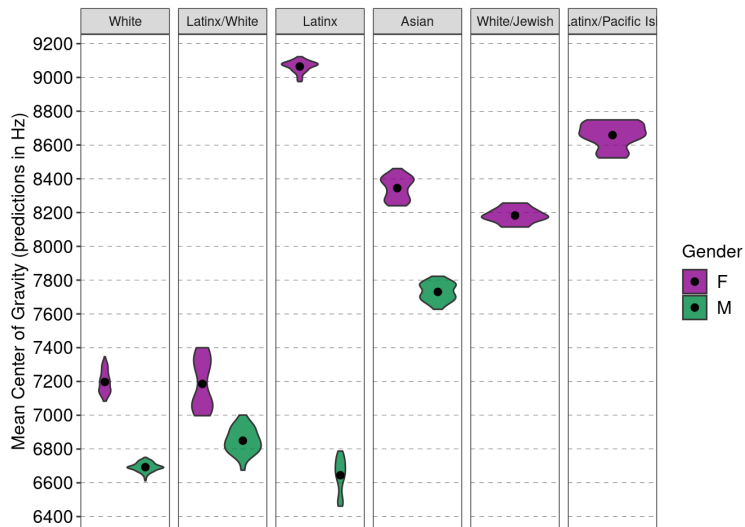


Figure 4. GENDER:RACE partial dependence co-plot

GENDER:MASCULINITY RATING was the second most important interaction in the COG random forest. Predictions generated from this sample situate the estimates for women around 7.45 kHz - 7.9 kHz and between 6.45 kHz - 6.7 kHz for men (Figure 5). Among women, estimated

COG generally increases as masculinity rating increases. In other words, the more masculine a woman rates herself, the further her predictions are. As for men, estimated COG generally decreases as masculinity rating increases. Put differently, the more masculine a man rates himself, the further back his predictions are.

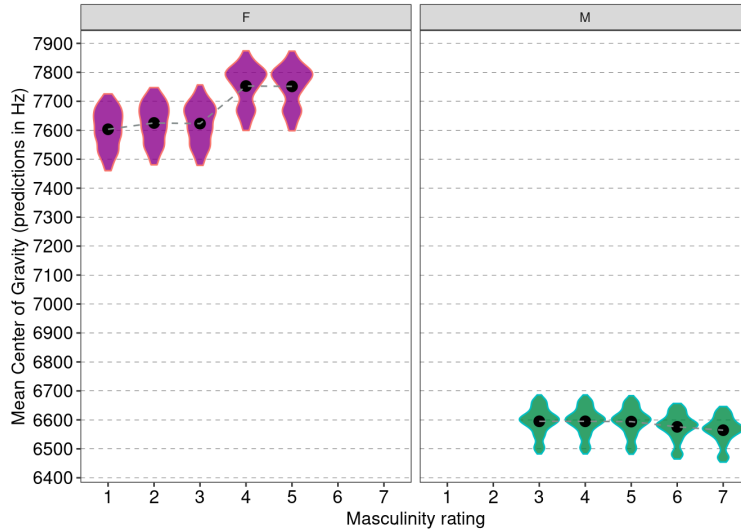


Figure 5. GENDER:MASCULINITY RATING partial dependence co-plot

GENDER:FEMININITY RATING is the next most important interaction in the COG random forest (Figure 5). Viewing women’s femininity ratings holistically, COG estimates generally increase as femininity ratings increase. That is, the more feminine a woman rates herself, the further her predictions are. However, COG estimates peak at a femininity rating of 5 out of 7 and then decreases slightly. Among men, there is a dramatic increase in estimated COG between men who rated themselves as not at all feminine (1) to a little feminine (2), and then another slight increase between a little (2) and moderately (3-4) feminine.

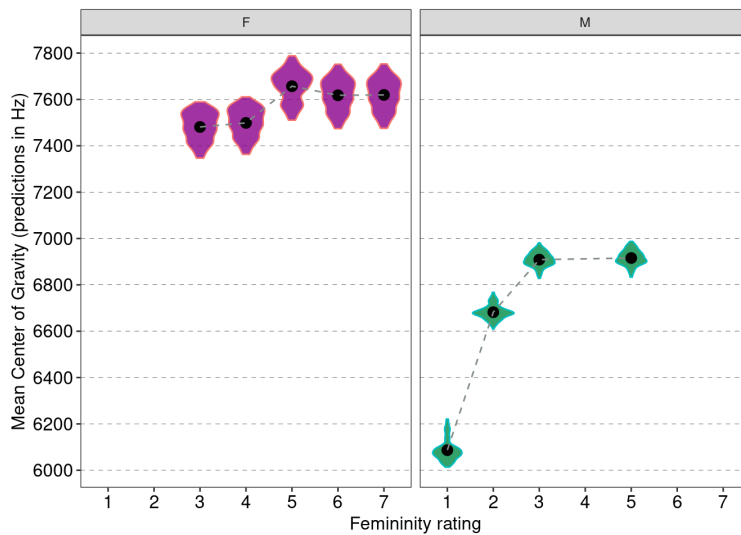


Figure 6. GENDER:FEMININITY RATING partial dependence co-plot

Finally, SEXUALITY:GENDER was also an important interaction in the COG random forest (Figure 6). Note that SEXUALITY as a single predictor is not important according to the VIMP values in Figure 2 above, but it participates in this important interaction so we discuss it here. Overall, straight women produce the highest COG estimates, followed by bisexual women, bisexual men, straight men, gay men, and finally lesbian women. Considering the within-sexuality, cross-gender comparisons, the difference between women and men’s predicted COG values is greatest among straight participants, followed by bisexual participants, and finally lesbian/gay participants. In the former two groups, women produce higher COG estimates than men. Unlike straight and bisexual speakers, however, lesbian women produce lower COG estimates than gay men.

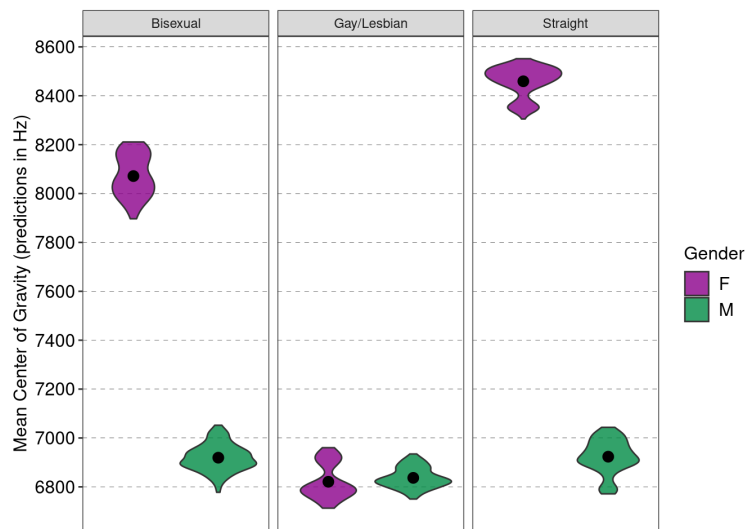


Figure 7. GENDER:SEXUALITY partial dependence co-plot

Turning to the within-gender, cross-sexuality comparisons, the differences between the three groups of men are rather minimal. In contrast, there is a visually salient difference between straight and bisexual women’s COG estimates compared to lesbian women’s, such that lesbian women produce substantially lower COG estimates than the other two groups of women. The difference between straight and bisexual women is also more extreme than the differences between the groups of men. In short, there seems to be much wider variation in predicted /s/ COG between the three groups of women than the groups of men in this sample.

6. Discussion. The results demonstrate the efficacy of random forests as a method for analyzing quantitative sociolinguistic data. The application of random forests to this dataset replicates and enriches previous findings (Willis forthcoming) by incorporating a broader set of predictors. Such an analysis is possible because random forests are well-equipped to analyze small, unbalanced, and non-random data structures with many dimensions of interest. We encourage researchers to consider random forests as a methodological intervention for similar research questions. In the remainder of this section, we focus our discussion on SEXUALITY, RACE, and GENDER because of their salience in our analysis.

The results indicate that bisexual speakers do not reliably pattern with lesbian/gay speakers or straight speakers. These findings call into question approaches that exclude bisexuality or

group bisexual speakers with other participants. Conflating bisexual speakers with lesbian/gay speakers a priori is a relatively common practice in experimental production studies (e.g., Munson et al. 2006a,b; Munson & Babel 2007).⁴ Bisexual erasure or exclusion is also prevalent in perception studies that elicit sexuality judgements. Two of the most common elicitation methods, the Likert scale paradigm (e.g., Munson et al. 2006a) and the forced choice paradigm (e.g., Smyth et al. 2003), render bisexuality either ambiguous or unavailable, respectively (Willis forthcoming). The widespread and uncritical engagement with these research practices is indicative of the pervasiveness of hegemonic monosexist ideologies in experimental studies of sexuality and the voice in English. We recommend that future research operationalize bisexuality as a distinct category unless data exploration, local epistemologies of sexuality, and/or theorization justify otherwise.

Next, we turn our attention to race and gender. First, however, we discuss an important caveat. Namely, the way ethnoracial categories were elicited likely primed a particular way of thinking about race. Participants were presented a list of terms (e.g. Asian, Latinx, etc.) as well as a free response section. They were instructed to circle one or more of the provided categories, write their own answer, or elaborate in the free response section, or to combine multiple options. This operationalization of race as distinct, discrete categories may not reflect the way participants understand race or experience racialization in everyday life. Given that many of the participants are unknown to us outside the research context, we are unable to assess the extent to which the presented operationalization captures participants' understandings and experiences of race and racialization.

What can be said based on this study is that the elicited ethnoracial distinctions appear to be highly relevant for predicting variation in /s/ frontedness. The GENDER:RACE interaction (Figure 3) suggests two broad patterns. First, members of the same gender do not have similar estimates across ethnoracial categories in our sample. That is, the mean and range of COG estimates for a particular gender is not consistent across ethnoracial groups. For example, Latinx women produce much higher COG estimates than Asian women, who produce much higher estimates than white women, etc. Second, ethnoracial groups vary substantially in terms of the degree of difference between women and men. For example, the estimated mean COG differs almost 2.5 kHz between monoracial Latinx women and men, but only about 500 Hz between white women and men. Wider variability between some women and men but not others suggests that ethnoracial groups may achieve gendered distinctions in /s/ production through different means, if such a distinction exists at all. Regardless, these results demonstrate that previous research on white or racially unspecified speakers (which we elaborate on below) may not be easily generalizable across ethnoracial communities. Indeed, this conclusion is not a novel one. Calder (2021) demonstrates that Black speakers in Bakersfield, California do not employ /s/ variation the way that is expected by the white canon. In fact, the indexical field of gender-based variation is inextricable from whiteness and is not incorporated into performances of local Black identity. In short, generalizations about gender differences in /s/ production are not consistent across ethnoracial groups for reasons that are inaccessible without deeper ethnographic understanding. All of this to say, our intention with this analysis is to support future qualitative research that accounts for how /s/ variation is used to produce (bisexual) identities in different queer communities across race and gender.

⁴ See also Pierrehumbert et al. (2004) who couch their confluations in data exploration.

7. Conclusion. Throughout this paper, we have demonstrated the efficacy of random forests as a possible methodological intervention to grapple with various difficulties many sociolinguists face in quantitative studies (see also Tagliamonte & Baayen 2012). Our random forest model included a large number of complex predictors related to identity categories, despite a number of potentially problematic characteristics of the data. However, the way in which these factors are quantified remains an issue. Eliciting the complex ways people experience identity and translating that complexity into discrete categories is no simple task. The way we elicited and operationalized ethnicity and race, for example, shaped the extent to which we could reasonably interpret our results. A one-size-fits-all solution to this issue is unlikely to exist. Instead, we argue that transparency in how and to what end participants' intersectional identities are elicited and operationalized is necessary. Previous work is not always forthcoming about aspects of their participants' identities that are not specifically under investigation. For example, studies typically do not report the ethnic or racial identities of their participants, much less include these categories as factors in any ensuing statistical modeling. When ethnicity/race is reported, it is usually to say that a small number of participants identify as non-white (Gaudio 1994), the implication being that speakers identify as white unless otherwise noted. Other studies make assumptions about participants' sexualities. (Smyth et al. 2003:334) report that some of their listeners "explicitly identified as gay males" whereas others "formed a mixed group, by which we mean that we did not ask about their sexual orientation and we presume that most identified as heterosexual". These practices raise questions about the generalizability of previous work. Comparison between studies is tenuous at best when there is limited information about the similarity of distinct groups of participants. Going forward, we urge researchers to be transparent about the various positionalities their participants occupy as well as how those identities were elicited and operationalized, even when those categories are not the focus of the analysis. Transparency and attention to research practice are vital steps towards combating the influence of monosexism and white privilege in studies of sexuality and the voice.

References

- Audacity Team. 2019. Audacity(R): Free audio editor and recorder. www.audacityteam.org.
- Barron-Lutzross, Auburn. 2015. The production and perception of a lesbian speech style. *UC Berkeley Phonology Lab Report* 11(11). <https://escholarship.org/uc/item/7f6332bh>.
- Ben Youssef, Chadi & Stefan Th Gries. forthcoming. Code-switching in Tunisian Arabic. *Corpora* 28.
- Breiman, Leo. 2001. Random forests. *Machine Learning* 45. 5–32. <https://doi.org/d8zjwq>.
- Calder, Jeremy. 2021. Whose indexical field is it?: The role of community epistemology in indexing social meaning. *Texas Linguistics Society* 39. 39–55.
- Campbell-Kibler, K. 2011. Intersecting variables and perceived sexual orientation in men. *American Speech* 86(1). 52–68. <https://doi.org/10.1215/00031283-1277510>.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4). 453–476. <https://doi.org/https://doi.org/10.1111/j.1467-9841.2008.00374.x>.
- Ehrlinger, John. 2016. ggRandomForests: Exploring random forest survival. <https://doi.org/j34g>.
- Fairbanks, Grant. 1960. *Voice and articulation drillbook*. New York: Harper & Row.
- Fuchs, Susanne & Martine Toda. 2010. Do differences in male versus female /s/ reflect biological or sociophonetic factors? In Susanne Fuchs, Martine Toda & Marzena Żygis (eds.),

- Turbulent sounds: An interdisciplinary guide*, 281–302. New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110226584.281>.
- Gaudio, Rudolf P. 1994. Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech* 69(1). 30. <https://doi.org/10.2307/455948>.
- Gries, Stefan Thomas. 2013. *Statistics for linguistics with R: A practical introduction* (2nd edn.) Berlin: De Gruyter Mouton.
- Hazenberg, Evan. 2015. Walking the straight and narrow: Linguistic choice and gendered presentation. *Gender and Language* 10(2). 270–294. <https://doi.org/10.1558/genl.v10i2.19812>.
- Ishwaran, Hemant. 2007. Variable importance in binary regression trees and forests. *Electronic Journal of Statistics* 1. 519–537. <https://doi.org/10.1214/07-EJS039>.
- Ishwaran, Hemant & Udaya B. Kogalur. 2022. randomForestSRC. R package.
- Linville, Sue Ellen. 1998. Acoustic correlates of perceived versus actual sexual orientation in men’s speech. *Folia Phoniatica et Logopaedica* 50(1). 35–48.
- Muchlinski, David, David Siroky, Jingrui He & Matthew Kocher. 2016. Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data. *Political Analysis* 24(1). 87–103. <https://doi.org/10.1093/pan/mpv024>.
- Munson, Benjamin. 2007. The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and Speech* 50. 125–142. <https://doi.org/d7kjsx>.
- Munson, Benjamin & Molly Babel. 2007. Loose lips and silver tongues, or, projecting sexual orientation through speech. *Language and Linguistics Compass* 1(5). 416–449. <https://doi.org/10.1111/j.1749-818X.2007.00028.x>.
- Munson, Benjamin, Elizabeth C. McDonald, Nancy L. DeBoe & Aubrey R. White. 2006a. The acoustic and perceptual bases of judgments of women and men’s sexual orientation from read speech. *Journal of Phonetics* 34(2). 202–240. <https://doi.org/10.1016/j.wocn.2005.05.003>.
- Munson, Benjamin, Sarah V. Jefferson & Elizabeth C. McDonald. 2006b. The influence of perceived sexual orientation on fricative identification. *The Journal of the Acoustical Society of America* 119(4). 2427–2437. <https://doi.org/10.1121/1.2173521>.
- Pierrehumbert, Janet B., Tessa Bent, Benjamin Munson, Ann R. Bradlow & J. Michael Bailey. 2004. The influence of sexual orientation on vowel production. *The Journal of the Acoustical Society of America* 116(4). 1905–1908. <https://doi.org/10.1121/1.1788729>.
- Podesva, Robert J. & Janneke Van Hofwegen. 2014. How conservatism and normative gender constrain variation in inland California: The case of /s/. *University of Pennsylvania Working Papers in Linguistics* 20(4).
- R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rogers, Henry & Ron Smyth. 2003. Phonetic differences between gay- and straight-sounding male speakers of North American English. *ICPhS* 15. 1855–1858
- Smyth, Ron, Greg Jacobs & Henry Rogers. 2003. Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society* 32(03). <https://doi.org/bsz3kr>.
- Strobl, Carolin, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4). 323–348. <https://doi.org/10.1037/a0016973>.
- Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178. <https://doi.org/10.1017/S0954394512000129>.

- Willis, Chloe. forthcoming. Bisexuality in experimental sociophonetics: Ideologies and implications. *Journal of Language and Sexuality* 13(1).
- Zimman, Lal. 2013. Hegemonic masculinity and the variability of gay-sounding speech. *Journal of Language and Sexuality* (2:1). 1–39.
- Zimman, Lal. 2017. Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants. *Linguistics* 55(5). <https://doi.org/gqhrtq>.
- Zimman, Lal. 2018. Duration & moments for /s/. Praat script.