CAMBRIDGE
UNIVERSITY PRESS

# TH-stopping in Philadelphia Puerto Rican English

Abigail E. Patchell[1] and Grant M. Berry[2]

[1]Department of Psychological and Brain Sciences, Villanova University, Villanova, PA, USA and
[2]Department of Spanish, Villanova University, Villanova, PA, USA
**Corresponding author:** Grant M. Berry. Email: grant.berry@villanova.edu

**Abstract**

Analyzing data from the Puerto Rican English in Philadelphia (PREP) corpus, we investigate participation in TH-stopping, a socially stigmatized yet stable variable documented in Philadelphia. While previous studies have been impressionistic and have considered voiced and voiceless tokens to pattern together, this work validates novel, acoustically based stopping indices: mean harmonics-to-noise ratio for voiced tokens and skewness for voiceless tokens. We apply these indices to the corpus data and analyze stopping under a Bayesian framework, and we compare results from a model built from impressionistic coding of a subset of the same data. We find convergent evidence that TH-stopping is a stable variable in the Puerto Rican English data as well. Findings are compared with those of existing studies, noting future directions for research on the variable and underscoring the importance of establishing demographically representative baselines for linguistic research in diverse urban centers.

## Introduction

In Philadelphia—one of the best-studied cities in the world with respect to language change—the lion's share of research has focused on White speakers. Although the variety commonly spoken by White speakers is locally prestigious and forms the basis for what is denoted "Philadelphia English" in scholarship and media (e.g., Gylfadottir, 2015; Labov, Rosenfelder, & Fruehwald, 2013; Purse, 2020; Sneller, Fruehwald, & Yang, 2019; Zellou & Tamminga, 2014), White non-Hispanics constitute only a third of the local population (U.S. Census Bureau, 2021). If the speech community rather than the individual is the basic unit of sociolinguistic investigation (e.g., Weinreich, Labov, & Herzog, 1968), it is imperative that the data accurately represent that community. Without a more representative picture of language from other socioeconomic and racioethnic groups in a community, we cannot accurately describe the types of change underway in that community, and we cannot fully understand how minority

communities contribute to patterns of language variation and change in the speech community more generally.

A primary goal of the current study is to complement existing work on Philadelphia English with data from a longstanding minority community: Philadelphia Puerto Ricans. Our reasons for selecting this population are twofold: (1) to expand the limited research on this underrepresented yet sizable subpopulation in Philadelphia; and (2) because members of this community have historically lived on the margins between predominantly White and predominantly Black neighborhoods, giving them continued exposure to speech from the other major ethnic groups in the city. Data collected from speakers in this community will permit us to investigate the extent to which features of a minority population reflect patterns observed in the surrounding community, here operationalized by the socially stigmatized phonological variable known as TH-stopping (where dental fricatives /θ, ð/ variably surface as stops [t, d]; Thomas, 2007).

This paper is organized as follows. First, we provide a broad overview of the history and current social reality of Puerto Ricans relative to other groups in Philadelphia. Next, we discuss studies describing the relationship between majority and minority language use, focusing on those that highlight Puerto Rican English elsewhere in the United States. We then describe previous research on the variable of interest (TH-stopping) and use controlled laboratory data to validate acoustic metrics for automatically coding the variable at scale. The validated metrics are then applied to data from a conversational corpus of spoken English from Puerto Ricans living in Philadelphia (the Puerto Rican English in Philadelphia [PREP] corpus), analyzing trends by Birth Year and Sex with Bayesian generalized additive mixed modeling. Next, we compare our findings with those from an analysis of more traditional impressionistic coding on a random subset of the corpus data administered to 23 raters via an online experimental paradigm. Finally, we situate the results alongside work that has been conducted on sound change in Philadelphia over the last century and work describing minority adoption of local patterns of language variation more generally.

### Community background

Puerto Ricans emerged as a prominent group in Philadelphia shortly after the turn of the twentieth century. They formed bonds with city merchants, providing them contacts and means in travel from Puerto Rico to Philadelphia. The group's presence increased during World War I, as more Puerto Ricans migrated to Philadelphia hoping to fill positions left open by labor shortages (Whalen, 2001). Migration continued after the Great Depression, with many post-Depression Puerto Ricans coming to work on farms in New Jersey and Pennsylvania. Unfavorable living conditions and poor wages eventually drove Puerto Ricans from farms into surrounding cities, particularly New York and Philadelphia. Shortly thereafter, Puerto Rican residents were joined by an influx of family, friends, and acquaintances from Puerto Rico, often spurred by the desire to find manufacturing jobs. Many Puerto Ricans found work in the garment industry, making a new home in North Philadelphia and Kensington (Ribeiro, 2022).

As Puerto Ricans migrated to mainland cities, Puerto Rican English began to take shape as its own structured variety. Shousterman (2014:30) described Puerto Rican

English as "a nonstandard variety of American English characteristically spoken by native English speakers in the United States who identify as being of Puerto Rican descent." Speakers of Puerto Rican English may be from different age-groups, but the variety is thought to have been shaped by the Puerto Rican migrants and their second-generation Puerto Rican children (many of whom acquired English natively). A commonly cited feature of Puerto Rican English is a syllable-timed prosodic rhythm, which has been linked to identity in older Puerto Ricans (Shousterman, 2014). However, in younger Puerto Ricans who live in close proximity to African Americans, this feature is less consistent. Puerto Rican English, like many nonstandard varieties, exhibits influence from speakers' interactions with members of the surrounding communities.

As the community continued to grow after World War II and into the 1970s, Philadelphia became the third largest urban Puerto Rican community in the United States (Casellas, 2007). Today, Philadelphia has the second-largest Puerto Rican community in the mainland US, behind only New York City (Ribeiro, 2022). Despite their prominence, Puerto Ricans in Philadelphia have been the subject of continued discrimination, racial tension, and hate crimes (Arnau, 2012). Interracial hostilities early on during their settlement in Philadelphia may have contributed to Puerto Ricans' tendency to live segregated from other racial groups—even as they gained political power and stronger representation in local government. Nonetheless, Puerto Ricans in Philadelphia have resided on the borders of both White and Black communities since their emergence as a prominent minority community (Ericksen, 1985), so their historical geographic segregation provides the potential for stratified exposure to distinct ways of speaking in the present.

### Sociodemographic characteristics of Philadelphia's Puerto Rican community

Language use in Philadelphia is starkly divided along racial and ethnic lines (e.g., Labov, 2014; Sneller, 2020; Wagner, 2014). By living on the margins between predominantly Black and predominantly White communities, many Philadelphia Puerto Ricans are exposed to two distinct varieties of English in their daily interactions. In the current study, we refer to these two varieties as White Philadelphia English (WPE) and Black Philadelphia English (BPE).[1] WPE (cf. Berry, 2018; Labov et al., 2013) generalizes speech patterns observed primarily in White speakers in Philadelphia and major sound changes over the last century. Currently, WPE is a variety characterized by an amalgamation of several traditionally Northern features (e.g., Canadian Raising: /aɪ/→[ʌɪ]) and traditionally Southern features (e.g., glide deletion before resonants /m,n,ɹ,l/; see also Labov et al., 2013). In contrast, BPE (cf. Fisher, 2022; Labov, 2014; Poplack, 1978; Sneller, 2020), which generalizes features observed in the speech of Black Philadelphians, is one of the many varieties of African American English (AAE). Other features of BPE, though, have largely been defined by their opposition to the same features in WPE. For example, while fronting of /aʊ/ as in "south" is undergoing reversal in WPE, this trend is not observed in BPE (Labov, 2014:5–6).

Beyond linguistic differences, WPE and BPE differ in the social implications of their use. WPE is a more prestigious variety, but it is also somewhat removed from the daily lives of many Philadelphia Puerto Ricans. Within the greater metropolitan area, Puerto

Ricans have less social mobility than White speakers, and there is a notable gap in mean socioeconomic status between the two groups. For example, 35% of Philadelphia Puerto Ricans currently fall below the poverty level, compared to 28% of Blacks and 12% of Whites who are not Hispanic (U.S. Census Bureau, 2020a, 2020b, 2020c). Philadelphia Puerto Rican and Black communities, in addition to being more likely to interact, have a social reality more like one another than the surrounding White community. Closer and more frequent connections to BPE speakers may serve as a catalyst for Puerto Rican adoption of BPE features, even though the majority variety could provide more opportunities for social mobility in the greater Philadelphia community. The speech patterns employed by Puerto Ricans in Philadelphia are likely complex, drawing at least in part from both surrounding varieties (although it is also possible that they exhibit features found in neither). In cases where both varieties exhibit a given linguistic feature, we would expect the Puerto Rican community to exhibit that feature as well. In cases where a feature exists in solely one of BPE or WPE, Puerto Rican English is likely to be influenced by the level of prestige (be it overt or covert) that the corresponding variety provides.

### Puerto Rican adoption of phonological variables

Poplack's (1978) study of Puerto Rican middle schoolers in Philadelphia explored teenagers' adoption of traditionally BPE and traditionally WPE variables and remains one of the only studies on Philadelphia Puerto Ricans to date. By manipulating speech style and analyzing discussions with the students, Poplack found that local prestige drove feature adoption: casual speech of young males showed a preference for BPE variables due to social solidarity with a Black peer, while females were more likely in all cases to use WPE features than BPE features. Moreover, although both groups showed style shifting, with increased WPE feature usage and decreased BPE feature usage in careful speech compared to casual speech, girls showed a greater distinction in usage rates between WPE and BPE variables in careful speech than boys did. The findings emphasize the importance of social perception on the adoption of minority speech patterns and suggest that those children, once adults, will continue to weigh social cues heavily in their adoption of BPE features versus WPE features.

   With respect to majority speech patterns, Berry (2018) found that adult Puerto Ricans, many of whom were born around the same time as Poplack's (1978) participants, were adopting three sound changes in-progress found in WPE: EY-raising (/eɪ/→[iː] in closed syllables), Canadian Raising (/aɪ/→[ʌɪ] preceding voiceless codas), and OH-lowering ({/ɔ/→[u]}→[ɒ]). These variables represent three points on the typical trajectory of language change. EY-raising, an indicator, shows social stratification but no apparent stylistic stratification and is under the level of social awareness. Canadian Raising, a marker, exhibits both social and stylistic stratifications, and the variable has spread to all sociodemographic groups. Finally, OH-lowering is a reversal of the highly stigmatized raised back vowel in Philadelphia (whence *wooder* "water"). As is the case for White Philadelphians (e.g., Labov et al., 2013), Berry (2018:56–62) found that younger Puerto Ricans were more likely to lower this vowel than older speakers, but the phonological realization was stratified by sex: gradual adoption

was found in females and adoption of the COT-CAUGHT merger was found in younger males. Berry (2018) also found little evidence of substrate effects of Spanish in Philadelphia Puerto Rican English.

While we know of no other studies that have attempted to describe linguistic features of Puerto Ricans in Philadelphia per se, Wolfram's (1971) study of Puerto Rican contact with speakers of other ethnicities in Harlem may provide insight into the effects of surrounding dialects. In this study, Wolfram focused on /d/ deletion and TH-fronting (/θ/→[f] / __+) to see if Puerto Ricans were adopting these features from the local Black community. Puerto Ricans with extensive Black contacts did indeed adopt more AAE features, but even speakers with few Black contacts adopted some AAE features, suggesting that feature adoption is possible even with limited social contact. This observation was replicated in a subsequent study, where Wolfram (1974) sampled conversational interviews conducted with both Black and Puerto Rican speakers to determine the prevalence of another variable related to fricative production: TH-stopping (/θ/→[t]; /ð/→[d]). Analysis of voiceless tokens suggested that the Puerto Ricans in NYC exhibited TH-stopping at higher rates than the Black speakers in the study. This finding was found to be unrelated to the individuals' extent of Black contacts, again illustrating that feature adoption can occur in cases of limited social contact.

### The variable under study: TH-stopping

TH-stopping has been a feature of working-class varieties of US English since at least the early twentieth century. Labov (2001) analyzed data from the Project on Language Change and Variation (LCV) in Philadelphia, finding that TH-stopping was a stable variable: while a curvilinear pattern for stopping rate was observed by socioeconomic status, there was no evidence of incrementation of the feature in apparent time (2001:119). It is also noted that voiced and voiceless word-initial tokens patterned similarly in NYC and, for all intents and purposes, can be considered a single variable (Labov, 2001:83). Importantly, though, most of the data from LCV come from White speakers (Labov, 2001:55), raising the questions of whether the variable is also stable and subject to the same constraints in other subsets of the community.

AAE speakers across the United States are also strongly associated with TH-stopping (e.g., Wolfram, 1974:66). Labov, Cohen, Robins, and Lewis (1968:170–171) observed an inverse correlation between frequency of stopping and both social class and spoken register: individuals of lower socioeconomic status were more likely to exhibit stopping, and those of a higher socioeconomic status remained aware of the stigma this holds. In research on middle-class Black speakers, Weldon (2021:131) postulated that TH-stopping was "a feature that gets perceived by mainstream listeners as indicative of lower social status, even if listeners are not able to overtly articulate what they are attending to linguistically." Despite this, Weldon (2021) noted that TH-stopping represented a "high risk, high reward" scenario: speakers are stigmatized for using the feature, but its strong association with Black identity provides covert prestige vis-à-vis solidarity with other speakers of the variety. Consequently, rates of stopping were much higher in conversational speech (81.5%) than formal speech (0.9%), and rates were higher when the speakers were familiar with one another than when they were not (Weldon, 2021:132).

### Current study

The data in the current study come from the PREP corpus (Berry, 2022). We used data available in a pre-release version of this corpus, which contains approximately 15–30 minutes of transcribed speech from each of 32 speakers. The speakers represented in these data were minimally second-generation Puerto Ricans and most ($n = 23$) have lived in Philadelphia their entire lives. No participants in the dataset reported preferring Spanish as a language of discourse. While we cannot ignore the possibility that Spanish has influenced their English, their longevity in an English-dominant locale and lack of preference for Spanish, when combined with previous research that has found little evidence of substrate effects from Spanish (Berry, 2018), casts doubt on language transfer as at least a primary source of TH-stopping. More information regarding the corpus and results from preliminary analyses can be found on the project webpage (dx.doi.org/10.17605/OSF.IO/7KM4R).

Since both BPE and WPE exhibit TH-stopping, we expect the variable will be adopted by Philadelphia Puerto Ricans as well. If so, the data would provide supplemental evidence that Puerto Rican speakers in large urban centers are likely to adopt features of other populations in their speech, and it would add to the limited body of research on this speech community. Another possibility is that TH-stopping is a change in-progress in the community. If speakers are aware of the social value of this variable (e.g., Labov, 2001:273–274), as is common with stigmatized variables, we would expect younger women to use TH-stopping less than younger men (e.g., Labov, 2001:97). If listeners are unaware of the social value of this variable, or if the variable is covertly prestigious, we would expect to see the highest rates in younger females (e.g., Eckert, 1989). It is also possible, though implausible, that Puerto Ricans in Philadelphia have not adopted TH-stopping at all. Based on existing research showing Puerto Rican adoption of local and supralocal patterns of language variation, we hypothesize that Puerto Ricans will adopt TH-stopping as a stable variable, as has been documented for other varieties of English in Philadelphia.

### Methods

### Toward a continuous measure for TH-stopping

To date, sociolinguistic studies of TH-stopping (and most other phonological variables) have been coded impressionistically: a trained individual listens to each token and determines whether the phone is realized as a stop, fricative, or—in some cases—an intermediate realization (e.g., Thomas, 2007). This approach suffers several drawbacks. First, a categorical coding schema may mask subtle patterns of variation present in the data. While we know of no studies that have demonstrated this for TH-stopping (but see Mitterer and Ernestus [2006], where spectral information was used to aid impressionistic coding), several studies on /s/ aspiration—a sociolinguistic variable widely researched in Spanish—have demonstrated that an acoustic-based analysis of a linguistic variable provided more granular data regarding the influence of social and linguistic factors on the acoustic realization than categorical approaches (Erker, 2010; File-Muriel & Brown, 2011; Ryant & Liberman, 2017). Second, impressionistic coding relies entirely on an outside listener and is not typically normalized

with respect to the speech patterns of the speaker. That is, impressionistic coding data cannot easily compare the similarity of a stop segment realized from an underlying fricative (e.g., *thin* /θɪn/ [tʰĩn]) to a stop segment that is underlyingly a stop (e.g., *tin* /tɪn/ [tʰĩn]). By taking an approach grounded in continuous acoustic metrics, we can directly compare how stop-like a phone is relative to unambiguous stops from the same speaker, which enables us to better understand the degree to which speakers distinguish the categories in production.

Another drawback of impressionistic coding is a practical one: manual coding is time consuming and not scalable to larger datasets without considerable cost and effort. While machine learning may be useful for automating classification problems and has been advocated as a promising frontier in this area (see, e.g., Villarreal, Clark, Hay, & Watson, 2020), such an approach does not easily overcome the disadvantages that may be inherent in categorical classification of gradient sociolinguistic variables to begin with. To identify potential acoustic correlates of TH-stopping, we followed Zhao (2010) by testing the reliability of metrics pertaining to energy distribution and periodicity on a controlled test set.

To begin, we created a small test set of 100 words, 25 of each beginning with /θ/, /ð/, /d/, or /t/; the full test set is available in the Supplementary Materials (https://osf.io/7km4r/). Two individuals were recorded reading three repetitions of the lists aloud in a sound-attenuated setting. We used *FAVE* (Rosenfelder, Fruehwald, Evanini, Seyfarth, Gorman, Prichard, & Yuan, 2022) to align the transcript to the audio data. We then wrote a *PRAAT* script to identify words with initial TH and extract metrics pertaining to energy distribution (skewness, kurtosis, center of gravity) and periodicity (mean harmonics-to-noise ratio [HNR]) over the consonant preceding and following phonetic context. Given that the aligned output frequently has leading spaces for word-initial phonemes, we set the measurement point 75% into the marked interval for all metrics. The script can be viewed in the Supplementary Materials.

We do not have an a priori reason to assume that voiced and voiceless tokens will behave similarly; it is possible that the distinct acoustic signatures of voiced versus voiceless tokens will result in differing correlations with the metrics analyzed. Moreover, we do not assume that the distribution of voiced and voiceless tokens within the speech community will be equal. In fact, distinct social embeddings of voiced and voiceless tokens may reveal that the adoption of stopping initially favors one voicing category over the other. Bearing this in mind, we separated the data by voicing and conducted separate analyses to determine the best predictors for each contrast.

Figure 1[2] shows the distribution of the test set data according to the contrasts investigated; standard errors are indicated by point ranges within the violin plots. Visual inspection suggests that mean HNR may reliably distinguish voiced tokens (/d/ versus /ð/), while skewness distinguishes voiceless tokens (/t/ versus /θ/).

To validate the visual trends and identify reliable indices of stopping, we fit two Bayesian mixed effects logistic regression models (one to the voiced tokens and another to the voiceless tokens) based on a binomial distribution in *R* (version 4.3.1; R Core Team, 2023) using the *rstanarm* package (version 2.21.4; Goodrich, Gabry, Ali, & Brilleman, 2023). The dependent variable was the underlying initial phone (/d/ versus /ð/, /t/ versus /θ/), and we included mean HNR, skewness, and kurtosis, as
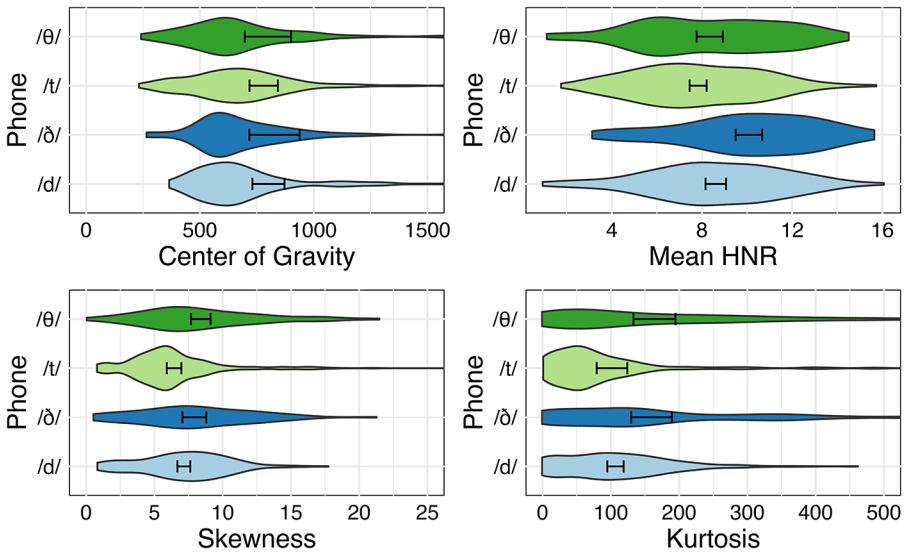
**Figure 1.** Distribution of test set tokens by phone and candidate metrics.

independent, fixed variables.[3] We included random intercepts by speaker to account for physiological and stylistic differences of the model talkers, and we used weakly informative priors because we did not have strong hypotheses regarding the influence of the metrics on classification. When fitting the model, we used a No-U-Turn Sampler (NUTS), a variant of Monte Carlo sampling, and set the *adapt_delta* parameter, which indicates the probability of accepting a given posterior draw, to a conservative value 0.99 (higher values more strongly penalize erratic jumps with high curvature between sampling steps). We sampled eight chains of 8,000 iterations from the posterior distribution, of which 2,000 were used for warmup and exploring the parameter space. We validated the model output using posterior predictive checks, leave-one-out (loo) validation with a Pareto-k diagnostic threshold of 0.7, and 95% highest density intervals (HDI) from the *hdi()* function in the *bayestestR* package (version 0.13.1; Makowski, Ben-Shachar, & Lüdecke, 2019). As a heuristic for assessing the importance of a parameter, we calculated the Bayes factor (BF) with the *bayesfactor()* function from the *bayestestR* package.

Results indicated the models were appropriately fit to the data without overgeneralizing. Pareto *k*-values were below 0.7, and posterior predictive checks were within acceptable ranges. First, we replicated visual trends: skewness ($CI_{95\%} = [-2.68, -1.14]$; BF = 3,450) reliably distinguished voiceless tokens while mean HNR ($CI_{95\%} = [-1.69, -0.58]$; BF = 399.46) distinguished voiced tokens (see also Figure 2). Beyond these, we found that center of gravity ($CI_{95\%} = [-1.23, -0.43]$; BF = 138.47) and kurtosis ($CI_{95\%} = [0.26, 1.61]$; BF = 5.55) predicted voiceless tokens. For voiced tokens, we found that skewness ($CI_{95\%} = [0.43, 2.33]$; BF = 16.11) and kurtosis ($CI_{95\%} = [-2.27, -0.57]$; BF = 48.05) were reliable predictors in addition to mean HNR. For the purposes of this study, we selected mean HNR as an indicator of voiced stopping and skewness as an indicator of voiceless stopping, as these matched
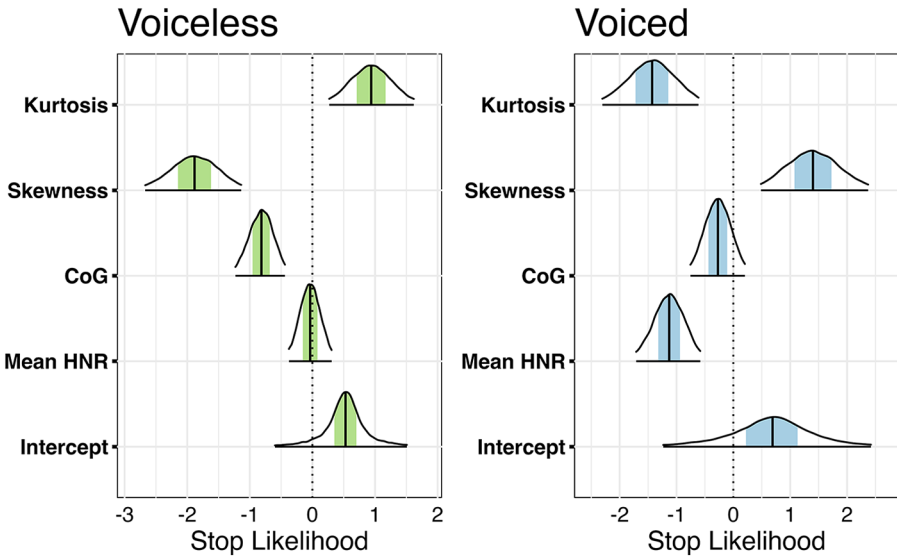
**Figure 2.** Posterior draws from GLMM by Metric and Voicing for test set data.

visual trends in both the raw data and the statistical model, and had considerably higher BFs than the other parameters.

### Token extraction and normalization

We used the aligned output from *FAVE* (Rosenfelder et al., 2022) to identify tokens with word-initial /ð/ ($n = 4982$) and /θ/[4] ($n = 596$) from the PREP corpus. We excluded any consonant clusters[5] (e.g., *three* /θɹiː/; $n = 174$) because coarticulation with the liquid in the consonant cluster could inflate the degree of stopping (e.g., O'Connor, 2001; Read, 1973; Resnick, 1972). This reduced the dataset to 4,982 voiced tokens and 422 voiceless tokens for analysis. Acoustic metrics were extracted with the same *PRAAT* script used for the test set.

### Outcome variables: Stopping indices

While we established that mean HNR and skewness are useful metrics for distinguishing fricatives from stops for voiced and voiceless tokens, respectively, the continuous metric does not provide a clear decision boundary (i.e., a threshold at which the fricative should be considered a stop). We thus frame the problem as one of distinction: how distinct are participants' underlying word-initial fricatives from their word-initial alveolar stops? To examine this, we extracted all words with an initial /d/ or /t/ from the PREP corpus, again excluding word-initial consonant clusters ($n = 443$). This left 1,805 /d/-initial words and 2,420 /t/-initial words for comparison with fricative tokens. Acoustic metrics were extracted with the same *PRAAT* script used for the fricative tokens. We required a minimum of five tokens per speaker and context for inclusion in the dataset.

Next, we created a speaker-normalized stopping index by dividing each token's distinguishing metric (mean HNR for voiced tokens; skewness for voiceless tokens) by the speaker's mean value for that metric in a similar phonetic context (i.e., with the same following vowel and stress). We then created centered variables representing a deviance from 1, which we used as dependent variables for statistical analysis. According to these deviance metrics, a value of zero means that an individual fricative token does not deviate from the same speaker's average unambiguous stop in the same phonetic context. To avoid extreme outliers, we also trimmed the data to exclude values three or more standard deviations beyond the group mean. This process removed 19% of the voiceless data ($n = 76$) and 23.3% of the voiced data ($n = 1163$). After removing outliers, 3,819 voiced tokens and 346 voiceless tokens were available for analysis. The 10 most frequent words for each underlying fricative in the trimmed dataset are included in Table 1. We note that only 2 of the top 10 words overall (THINK, THINGS) began with a voiceless fricative.

**Table 1.** Most frequent words by Voicing

| /ð/ | | /θ/ | |
|---|---|---|---|
| Word | *n* | Word | *n* |
| the | 1324 | think | 101 |
| that | 717 | things | 82 |
| they | 576 | thing | 72 |
| then | 223 | thousand | 21 |
| there | 215 | thought | 11 |
| that's | 201 | thank | 10 |
| this | 200 | thinking | 8 |
| them | 104 | therapy | 6 |
| they're | 52 | thyroid | 6 |
| those | 43 | thirty | 5 |

### *Impressionistic coding of a random sample of data*

While the speaker-normalized metrics are based on statistical differences in the acoustic values, it is possible that those differences do not reflect the same differences picked up by listeners. Likewise, it is possible that the reliability of the metrics for distinguishing fricatives from stops shown in the test set data does not generalize to less controlled conversational data. Given these possibilities and the primacy of impressionistic coding in the existing literature on TH-stopping, we selected a random subset of up to 25 tokens by phone per speaker ($n = 1020$)[6] to be coded impressionistically. We created a categorization task in *OpenSesame* (Mathôt & March, 2022; Mathôt, Schreij, & Theeuwes, 2012), which we uploaded to a JATOS server and administered via Amazon Mechanical Turk to 23 expert workers located in the United States. In the task, raters first calibrated their audio to ensure they could hear the stimuli at a comfortable volume. Then, they were introduced to the experiment with the following text:

In some varieties of English, the TH sounds in words can be pronounced more like "t" or "d" sounds. For example, THAT might sound more like DAT and THINK might sound more like TINK. Your goal in this study is to determine whether the sound is more of a fricative (e.g., "that") or more of a stop (e.g., "dat").

On the following screens, you will see a keyword. Please listen for that keyword in the speech sample. If the TH sound in the keyword sounds more like a stop ("tin" or "dis"), press the S key. If the TH sound in the keyword sounds more like a fricative ("thin" or "this"), press the K key. If you can't determine what the sound is, or if you aren't sure, please respond with the space bar.

Once participants confirmed they understood the expectations of the task, they were shown the target word on the screen and presented with a sample of audio containing the target word. Participants indicated via the keyboard whether the initial sound in the target word appeared to be a stop, a fricative, or indeterminate (see Figure 3 for a sample response screen). There was no timeout for an individual trial, and participants were given a short break after every 100 trials. The task took approximately 30 minutes to complete, and participants were compensated $5 for their time.
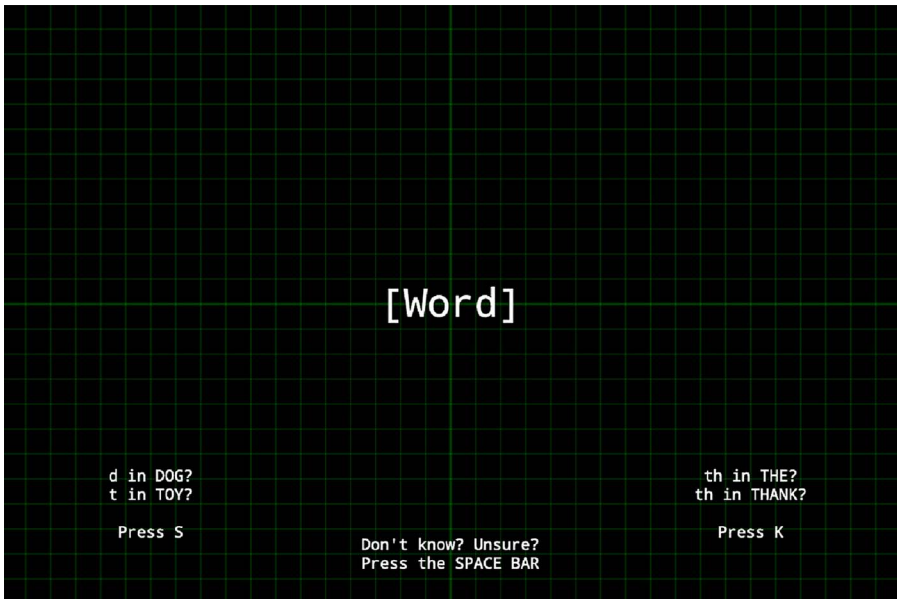


**Figure 3.** Sample response screen from impressionistic coding experiment.

### Analysis: Bayesian generalized additive mixed models

To investigate the influence of Birth Year and Speaker Sex on production patterns, we created generalized additive mixed models using the *stan_gamm4* function in the *rstanarm* package (version 2.21.4; Goodrich et al., 2023). We used treatment coding for

Sex, setting the reference level at Male. Fixed effects included Birth Year, Sex, and their interaction. Additionally, we added smooth functions for Birth Year by Sex. To constrain these smooth terms and ward against overfitting, we constrained the model to detect only basic curvilinear trends and penalize excessive curvature by manipulating the *k* and *m* parameters for the smooth functions. The *k* parameter (number of knots) determines the dimension of the basis vector for the smooth function; a value of 3 permits at most a three-dimensional basis and in general only very basic curvilinear patterns. The *m* parameter is the order of the penalty applied to the smooth function; the value of 2 is standard for splines built on cubic polynomials (Wood, 2003). The random effects structure included random intercepts for Speaker to account for by-speaker variability.

Regarding general model parameters, we followed a procedure similar to that of the test set analysis. We set the *adapt_delta* parameter to 0.99 and sampled eight chains of 8,000 steps (2,000 warmup). We assessed model fit using posterior predictive checks and loo cross validation. We report the effective sample size for key parameters from the statistical model, which provide insight into correlation among predictors and confidence of the individual parameter estimate. We evaluated parameter estimates according to 95% HDI calculated using the *hdi()* function from the *bayestestR* package (version 0.12.1; Makowski et al., 2019) and, to interpret effects, we examined the percentage of posterior draws that fell within the region of practical equivalence (ROPE) using the *rope()* and *rope_range()* functions from the same package. Specifically, we follow Kruschke's (2018) HDI + ROPE decision rule: we accept the null hypothesis in cases where the entire 95% HDI falls within the ROPE, reject the null hypothesis when the entire 95% HDI falls outside the ROPE, and withhold a decision in all other cases.

## Results

### *Normalized stopping ratios*

Figure 4 plots generalized additive model smooths for the raw production data with the same shape parameters as the statistical models but without random intercepts by speaker.[7] Although there is considerable variation, men and women from all age-groups appear to show little distinction between fricatives and stops on average (evidenced by a deviance value of zero). This is echoed in the statistical models, which do incorporate individual variability in their random effects structures (see Table 2). Of note is that the 95% HDI for Birth Year and the interaction between Birth Year and Sex fall entirely within the ROPE for both the voiced and voiceless models; this provides strong evidence in support of the null hypothesis that TH-stopping is not a function of Birth Year, neither independently nor as a covariate with Sex. Figure 5 provides a nonlinear plot of the smooth by Birth Year and Sex for the voiceless and voiced models, which also reflects these findings: the degree of variability ranges widely across different age-groups, but, on average, stops and fricatives do not deviate from one another acoustically for the metrics analyzed (HNR for voiced tokens and skewness for voiceless tokens).

### *Impressionistic coding*

A cross-tabulation of stopping rates based on impressionistic coding is presented in Table 3. For ease of presentation, we divided speakers into two age-groups based
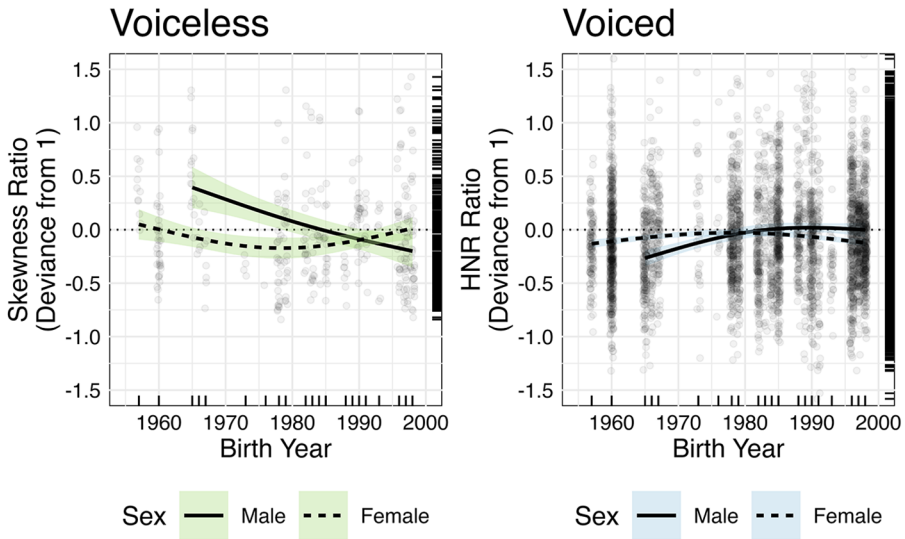
**Figure 4.** Normalized stopping indices by Sex, Birth Year, and Voicing.

**Table 2.** Summary of GAMMs from production data

| Parametric terms | Voiceless | | | Voiced | | |
|---|---|---|---|---|---|---|
| | HDI<sub>95%</sub> | % in ROPE | $n_{eff}$ | HDI<sub>95%</sub> | % in ROPE | $n_{eff}$ |
| Intercept | [−45.07, 68.84] | 0.15 | 10,047 | [−66.83, 53.56] | 0.19 | 10,059 |
| Sex (Female) | [−3.53, 3.45] | 2.17 | 20,838 | [−3.70, 4.00] | 2.15 | 21,484 |
| Birth Year | [−0.03, 0.02] | 100.00 | 10,046 | [−0.03, 0.03] | 100.00 | 10,058 |
| Sex: Birth Year | [0.0, 0.0] | 100.00 | 20,858 | [0.0, 0.0] | 100.00 | 21,443 |
| **Smooth terms** | **HDI<sub>95%</sub>** | **% in ROPE** | **$n_{eff}$** | **HDI<sub>95%</sub>** | **% in ROPE** | **$n_{eff}$** |
| s(BirthYear): SexMale1 | [0.00, 0.92] | 12.05 | 20,402 | [0.00, 1.09] | 9.03 | 12,413 |
| s(BirthYear): SexMale2 | [0.00, 0.86] | 12.14 | 10,350 | [0.00, 0.96] | 13.38 | 10,323 |
| s(BirthYear): SexFemale1 | [0.00, 1.01] | 6.72 | 10,682 | [0.00, 0.93] | 13.25 | 8,250 |
| s(BirthYear): SexFemale2 | [0.00, 0.85] | 13.03 | 10,220 | [0.00, 0.97] | 13.47 | 10,171 |

on their birth year relative to 1985 (see Berry, 2018 for a similar procedure). The raw data suggest that voiced fricatives are more likely to be stopped (or at least perceived as stopped) than voiceless tokens. Younger and Older speakers have similar stopping rates for voiced tokens, and men and women have similar stopping rates overall. It is worth
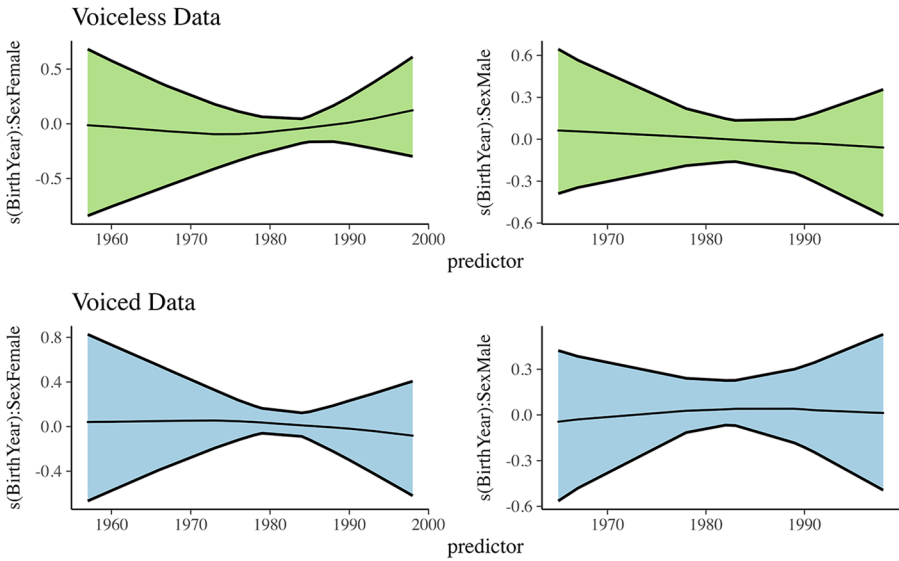
**Figure 5.** Smooth plots for Birth Year by Sex and Voicing for production data from GAMMs.

noting that the voiceless stopping rates for Puerto Ricans in Philadelphia in 2017 are slightly lower than the 25.9% reported for pre-vocalic, morpheme-initial TH for Puerto Ricans in New York in Wolfram's (1974:72) study. Though there is no exact data point in Philadelphia for stopping rate, Labov (2001:78) created a (dh) index for voiced stopping where fricative tokens were coded as 0, intermediate tokens were coded as 1, and stopped tokens were coded as 2. We generated a similar index, finding values of 52.7 for younger men compared to 45.7 for younger women. Both values fall on the low end of what Labov (2001:96) found for Philadelphians, which ranged from approximately 40 to 130. We will address possible reasons for these discrepancies in the *Discussion* section.

**Table 3.** Stopping rates by Age-group, Sex, and Voicing based on impressionistic coding

| Age-group | Sex | Speakers | Voiceless (/θ/) | | Voiced (/ð/) | | |
| | | | Stopping rate (%) | n | Stopping rate (%) | n | (dh) index[a] |
|---|---|---|---|---|---|---|---|
| Older (Born before 1985) | Male | 5 | 11.1 | 1436 | 23.8 | 1898 | 44.5 |
| | Female | 11 | 15.3 | 2694 | 24.1 | 4142 | 47.5 |
| Younger (Born after 1985) | Male | 6 | 15.0 | 1126 | 27.8 | 2280 | 52.7 |
| | Female | 10 | 15.5 | 2592 | 22.4 | 3793 | 45.7 |

[a]Approximating Labov (2001:78).

A plot of stopping rates based on raters' impressionistic codes by Sex, Voicing, and Birth Year with GAM smooths by Sex (Figure 6) echoes observations from the summary statistics in Table 3 (nonlinear plot provided in Figure 7). Men and women do
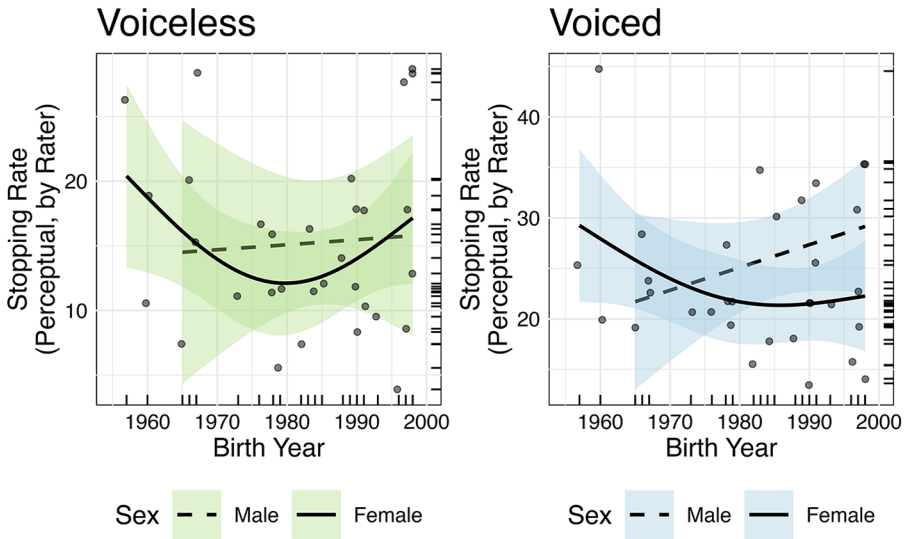
**Figure 6.** Stopping rates by Birth Year and Sex based on impressionistic coding.
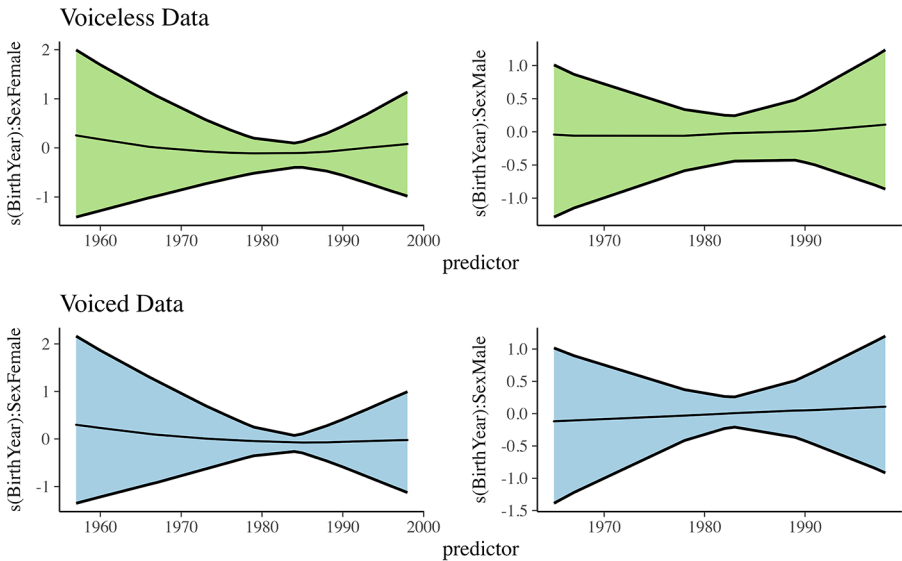


**Figure 7.** Smooth plots for Birth Year by Sex and Voicing for impressionistic data from GAMMs.

not appear to differ in perceived stopping rate, and voiced stopping rates appear higher than voiceless stopping rates overall. Finally, and most importantly, stopping rates are easily above zero for both voiced and voiceless tokens. Ergo, we can conclude that Puerto Ricans in Philadelphia are participating in TH-stopping, at least to some degree.

The output from the statistical models reflects these general trends (Table 4). As was the case with the models built on the raw production data, we find that the 95% HDIs for Birth Year and the interaction between Sex and Birth Year fall entirely within the ROPE. This provides statistical evidence to support the same conclusion found in the acoustically coded data: TH-Stopping is not influenced by Birth Year, neither independently nor as an interaction with Sex.

**Table 4.** Summary of GAMMs from impressionistic coding

| Parameter | Voiceless | | | Voiced | | |
|---|---|---|---|---|---|---|
| Parametric terms | $HDI_{95\%}$ | % in ROPE | $n_{eff}$ | $HDI_{95\%}$ | % in ROPE | $n_{eff}$ |
| Intercept | [−126.58, 132.76] | 0.43 | 9827 | [−134.37, 134.05] | 0.40 | 8196 |
| Sex (Female) | [−7.16, 7.45] | 4.09 | 19,162 | [−7.04, 7.53] | 4.22 | 16,582 |
| Birth Year | [−0.07, 0.06] | 100.00 | 9826 | [−0.07, 0.07] | 100.00 | 8195 |
| Sex: Birth Year | [−0.00, 0.00] | 100.00 | 19,182 | [−0.00, 0.00] | 100.00 | 16,602 |
| Smooth terms | $HDI_{95\%}$ | % in ROPE | $n_{eff}$ | $HDI_{95\%}$ | % in ROPE | $n_{eff}$ |
| s(BirthYear): SexMale1 | [0.00, 2.24] | 19.98 | 15,593 | [0.00, 1.94] | 26.54 | 12,795 |
| s(BirthYear): SexMale2 | [0.00, 1.95] | 26.58 | 10,293 | [0.00, 1.98] | 26.59 | 8416 |
| s(BirthYear): SexFemale1 | [0.00, 2.33] | 13.56 | 8224 | [0.00, 1.99] | 22.01 | 7048 |
| s(BirthYear): SexFemale2 | [0.00, 1.95] | 25.83 | 9740 | [0.00, 1.99] | 26.28 | 8160 |

## Discussion

Both analyses coincide in demonstrating that TH-stopping is a stable variable in Philadelphia Puerto Rican English. We find convergent evidence in support of the null hypothesis regarding Birth Year, so the data do not support an account of age grading or change in-progress. In contrast to Labov's (2001:97) analysis, we did not find strong evidence that women have lower rates of voiced stopping than men. However, we note that Sex as an independent factor had indeterminate results in both models and thus we cannot conclusively argue for or against an effect of Sex. That said, it is worth noting that very little of the ROPE for any of the analyses contained Sex (the maximum was 4.22% in the impressionistic model of voiced tokens); it is possible that a larger-scale study of the variable with more speakers may find conclusive evidence of an effect of Sex or lack thereof.

Lack of effects for both Sex and Birth Year could mean that stopping has become more socially acceptable within the Philadelphia speech community, or it may be the case that Philadelphia Puerto Ricans are unaware of its social stigma in the broader community (e.g., Santa Ana & Parodi, 1998). To confirm the latter, we would require a broader sample of Puerto Rican speech that would permit a detailed analysis of

the variable by socioeconomic status and speech style. To evaluate the former, a contemporary analysis reexamining usage patterns in Philadelphia's White and Black populations—the other two major ethnic groups in the city—is necessary. We can say, though, that males and females do not differ in the influence of age on their use of TH-stopping. Additionally, we can say that TH-stopping is a feature of Puerto Rican English in Philadelphia, as it is in other English varieties in Philadelphia.

One might be tempted to attribute some or all of these findings to transfer from Spanish given that (1) Spanish lacks a phonemic distinction between /d/ and /ð/ and instead treats both as allophones of the same underlying category and (2) that /θ/ is not phonemic in varieties of Spanish outside Central and Northern Spain. Though convenient, a language transfer account is unlikely. Speakers in the PREP corpus use English as their primary language and have done so for most of their lives. Moreover, there is little evidence elsewhere of phonological transfer from Spanish in their English production. For example, they maintain a tense/lax distinction for all vowels, including those that are notoriously difficult for native Spanish speakers learning English (e.g., /i/-/ɪ/; see Berry, 2018:24–26). Importantly, if transfer from Spanish were truly the underlying cause of stopping, we would have expected much higher rates of stopping in the impressionistic data than were found.

While our results indicate that TH-stopping is a stable variable in Philadelphia Puerto Rican English, the stopping rates found in this study are lower than have been previously reported. One possible explanation is that the novel metrics explored were unreliable, though they were validated with a controlled test set. It is also possible that there were issues with the training level of the raters, but there were 23 individual raters whose idiosyncrasies were controlled for statistically, and findings from analyses of their ratings aligned with the acoustically based analyses. Another explanation, perhaps most plausible based on existing data, is that TH-stopping is not as characteristic of this community as has been found for others in Philadelphia and the surrounding area.

### *Limitations and future directions*

Although our findings add to the limited body of research concerning sound change in Puerto Rican Philadelphians and appear to be internally valid, we acknowledge that this was a small-scale, introductory study. A more robust dataset is necessary to provide a more holistic picture of the community and its members' use of language. Nonetheless, we believe that despite the small sample size, the data still provide important insights which can be used to generate future hypotheses.

Like all corpus data, the PREP dataset reflects the distribution of words in the language. This means an abundance of function words, and many of these that begin with TH in English are underlyingly voiced (see Table 1). Given the high frequency of such words and the reducing effect of frequency (e.g., Gahl, 2008), function words may be more susceptible to stopping than other classes of words, so it is possible that the dataset inflates the degree of stopping. However, it is also likely that prior corpus analyses of voiced TH-stopping contain a similar distribution of words, so such a limitation is not unique to this study. Subsequent research analyzing stopping by phonetic context and general frequencies in that context (cf. Brown, 2018; Brown & Raymond, 2012) would

provide a fuller picture of the influence, if any, of the imbalance of content and function words across voicing groups on stopping.

Imbalanced stratification of speakers by socioeconomic status is a primary shortfall of this study. Speakers in the PREP corpus are skewed toward upper-working-class and lower-middle-class, and a third ($n = 13$) were unemployed at the time of data collection. While these social strata are precisely those where we would expect to first see changes from below, we cannot confirm this without comparing data from these groups to a full range of socioeconomic strata. To do this, additional data from the community will need to be collected and transcribed.

Additionally, the lower number of men included in the study's dataset limits its statistical power. Since increasing the number of men in the sample would likewise require additional data collection and is thus not in the scope of this study, a focus of future research should be to increase the sample size by prioritizing adding men to the corpus. Doing so would have the added benefit of clarifying whether there is a main effect of Sex on TH-stopping in the community or not.

A third limitation is that we do not quantify speakers' degree of contact with other ethnic groups in Philadelphia, so we cannot know exactly how interaction and attitudes influence their adoption of TH-stopping. One way to address this question would be to conduct a social network analysis approximating contact between Black and White Philadelphians with US Census data. Doing so would take steps toward determining how an individual's social structure in the Puerto Rican community in Philadelphia influences their adoption of the newer voiceless TH-stopping. A more holistic picture of these speakers' language environment could also be obtained from additional survey data asking about the kinds of people Puerto Rican Philadelphians usually interact with.

## Conclusion

This work expands the limited research on the Puerto Rican community and provides an important complement to analyses that have to-date focused on the speech of White Philadelphians. We introduced two novel, context-controlled, acoustically based metrics for assessing the degree of fricative stopping, and we compared these findings to those of more traditional impressionistic classification. Findings from acoustic and impressionistic analyses coincided, providing evidence for TH-stopping as a stable variable within the Philadelphia Puerto Rican speech community.

In a situation where speakers exist on the threshold between linguistically distinct communities, feature adoption is rarely unilateral. Speakers may selectively adopt features from those communities, or they may innovate internally. Flexibility in feature adoption is another rich source of variation that provides researchers with the opportunity to explore the influence of social reality on linguistic structure. Further research on communities such as these, then, is essential for understanding how speech communities are formed and how patterns of language use are established and spread.

preliminary version of this manuscript. The feedback from several anonymous reviewers has also greatly improved this manuscript. The authors assume responsibility for any remaining errors.

**Competing interests.** The authors declare none.

## Notes

**1.** In using the terms BPE and WPE, we do not intend to imply that speakers in each ethnic group are homogeneous in their use of language. Rather, these represent generalizations from observed trends in previous linguistic research stratified by ethnic group in the broader Philadelphia population.

**2.** An *R* Notebook and a rendered HTML page containing code and outputs from all analyses are available in the Supplementary Materials. These scripts also contain code that will launch an interactive app for model evaluation and visualization using *shinystan* (version 2.6.0; Gabry & Veen, 2022).

**3.** Our aim is not to discover the best possible metric (or combination thereof) for distinguishing stops and fricatives. Rather, we seek to identify a simple, phonetically plausible, and reliable metric that can be employed at scale. We encourage future researchers to iterate and improve on this approach.

**4.** *FAVE* uses the CMU pronouncing dictionary to determine the phones in each word, and this dictionary does not permit stopped variants of word-initial TH as alternative pronunciations. As such, there is no concern that alignment/segmentation by *FAVE* biases the token distribution in favor of more fricative-like (i.e., canonical) tokens.

**5.** In English, the only possible word-initial consonant cluster with an initial dental fricative is /θɹ/, so this exclusion only applied to voiceless tokens.

**6.** The randomly sampled tokens were saved in the format SPEAKER_WORD_START_END. Both the start and end points were rounded to the nearest second, which we realized a posteriori was insufficient to uniquely map each impressionistic token back to the original production token. As a consequence, a direct one-to-one comparison of impressionistic code and acoustic metric was not feasible.

**7.** *R* syntax: $\text{plot} + \text{geom\_smooth}(\text{formula} = y \sim s(x, k = 3, m = 2, \text{method} = \text{`gam'}))$.

## References

Arnau, Ariel. (2012). The evolution of leadership within the Puerto Rican community of Philadelphia, 1950s-1970s. *The Pennsylvania Magazine of History and Biography* 136(1):53–81.

Berry, Grant M. (2018). *Liminal Voices, Central Constraints: Minority Adoption of Majority Sound Change*. Doctoral dissertation, The Pennsylvania State University.

Berry, Grant M. (2022). Language variation and change in Puerto Rican Philadelphia. https://doi.org/10.17605/OSF.IO/7KM4R.

Brown, Esther L. (2018). Cumulative exposure to phonetic reducing environments marks the lexicon: Spanish /d-/ words spoken in isolation. In K. Aaron Smith & D. Nordquist (Eds.), *Functional and Usage-based Approached to the Study of Language: In Honor of Joan L. Bybee*. New York: John Benjamins. 127–153.

Brown, Esther L., & Raymond, William D. (2012). How discourse shapes the lexicon: Explaining the distribution of Spanish *f-/h-* words. *Diachronica* 29(2):139–161.

Casellas, Marisa. (2007). El Barrio: Latino relationships in North Philadelphia and impacts on Puerto Rican businesses. *Urban Studies Seminar Papers* 1:1–39.

Eckert, Penelope. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change* 1:245–267.

Ericksen, Eugene P. (1985). *The state of Puerto Rican Philadelphia*. Philadelphia: Institute for Public Policy Research.

Erker, Daniel. (2010). A subsegmental approach to coda /s/ weakening in Dominican Spanish. *International Journal of the Sociology of Language* 203:9–26.

File-Muriel, Richard J., & Brown, Earl K. (2011). The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change* 23:223–243.

Fisher, Sabriya. (2022). The status of *ain't* in Philadelphia African American English. *Language Variation and Change* 34:1–28.

Gabry, Jonah, & Veen, Duco (2022). shinystan: Interactive visual and numerical diagnostics and posterior analysis for Bayesian models. R package version 2.6.0.

Gahl, Susanne. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3):474–496.

Goodrich, Ben, Gabry, Jonah, Ali, Imad, & Brilleman, Sam. (2023). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.4.

Gylfadottir, Duna. (2015). Shtreets of Philadelphia: An acoustic study of /str/-retraction in a naturalistic speech corpus. *University of Penn Working Papers in Linguistics* 21(2):11.

Kruschke, John K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science* 1(2):270–280.

Labov, William. (2001). *Principles of Linguistic Change, Volume 2: Social Factors*. Oxford: Blackwell.

Labov, William. (2014). The role of African Americans in Philadelphia sound change. *Language Variation and Change* 26:1–20.

Labov, William, Cohen, Paul, Robins, Clarence, & Lewis, John. (1968). *A study of the non-standard English of Negro and Puerto Rican Speakers in New York City. Cooperative Research Report 3288. Vols I and II.* Philadelphia: U.S. Regional Survey (Linguistics Laboratory, University of PA).

Labov, William, Rosenfelder, Ingrid, & Fruehwald, Josef. (2013). One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. *Language* 89:30–65.

Makowski, Dominique, Ben-Shachar, Mattan S., & Lüdecke, Daniele. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software* 4(40):1541. R package version 0.12.1.

Mathôt, Sebastiaan, & March, Jennifer. (2022). Conducting linguistic experiments online with OpenSesame and OSWeb. *Language Learning* 72(4):1017–1048.

Mathôt, Sebastiaan, Schreij, Daniel, & Theeuwes, Jan. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44(2):314–324.

Mitterer, Holger, & Ernestus, Mirjam. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34:73–103.

O'Connor, Kathleen M. (2001). When is a cluster not a cluster? *Clinical Linguistics & Phonetics* 15(1–2):53–56.

Poplack, Shana. (1978). Dialect acquisition among Puerto Rican bilinguals. *Language in Society* 7:89–103.

Purse, Ruaridh. (2020). The envelope of variation for /l/ vocalization in Philadelphia. *University of Pennsylvania Working Papers in Linguistics* 25(2):12.

R Core Team. (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Version 4.3.1 'Beagle Scouts'.

Read, Charles. (1973). Children's judgment of phonetic similarities in relation to English spelling. *Language Learning* 23:17–38.

Resnick, Melvyn C. (1972). The redundant English phonemes /č,ǰ,š,ž/. *Linguistics* 10(86):83–86.

Ribeiro, Alyssa (2022). *Puerto Rican Migration*. Encyclopedia of Greater Philadelphia.

Rosenfelder, Ingrid, Fruehwald, Josef, Evanini, Keelan, Seyfarth, Scott, Gorman, Kyle, Prichard, Hilary, & Yuan, Jiahong. (2022). FAVE (Forced Alignment and Vowel Extraction) Program Suite v2.0.0/zenodo.

Ryant, Neville, & Liberman, Mark. (2017). Large-scale analysis of Spanish /s/-lenition using audiobooks. *Proceedings of Meetings on Acoustics* 28:060005.

Santa Ana, Otto, & Parodi, Claudia. (1998). Modeling the speech community: Configuration and variable types in the Mexican Spanish setting. *Language in Society* 27:23–51.

Shousterman, Cara. (2014). Speaking English in Spanish Harlem: The Role of Rhythm. *University of Pennsylvania Working Papers in Linguistics* 20:159–168.

Sneller, Betsy. (2020). Phonological rule spreading across hostile lines: (TH)-fronting in Grays Ferry. *Language Variation and Change* 32(1):25–47.

Sneller, Betsy, Fruehwald, Josef, & Yang, Charles. (2019). Using the Tolerance Principle to predict phonological change. *Language Variation and Change* 31(1):1–20.

Thomas, Erik. (2007). Phonological and phonetic characteristics of African American Vernacular English. *Language and Linguistics Compass* 1:450–475.

U.S. Census Bureau. (2020a). Table B17001I: Poverty status in the last 12 months by sex and age (Hispanic or Latino). 2020 ACS 5-year Estimates detailed tables.

U.S. Census Bureau. (2020b). Table B17001H: Poverty status in the last 12 months by sex and age (White alone, not Hispanic or Latino). 2020 ACS 5-year Estimates detailed tables.

U.S. Census Bureau. (2020c). Table B17001B: Poverty status in the last 12 months by sex and age (Black or African American alone). 2020 ACS 5-year Estimates detailed tables.

U.S. Census Bureau. (2021). QuickFacts. Philadelphia, PA. Available at https://www.census.gov/quickfacts/fact/table/philadelphiacitypennsylvania/PST045221.

Villarreal, Dan, Clark, Lynn, Hay, Jennifer, & Watson, Kevin. (2020). From categories to gradience: Auto-coding sociophonetic variation with random forests. *Laboratory Phonology* 11:1–31.

Wagner, Suzanne Evans. (2014). Linguistic correlates of Irish-American and Italian-American ethnicity in high school and beyond. *Language and Communication* 35:75–87.

Weinreich, Uriel, Labov, William, & Herzog, Marvin I. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann, and Yakov Malkiel (Eds.), *Directions for Historical Linguistics.* Austin: University of Texas Press. 95–195.

Weldon, Tracey. (2021). *Middle Class African American English*. New York: Cambridge University Press.

Whalen, Carmen Teresa. (2001). *From Puerto Rico to Philadelphia: Puerto Rican Workers and Postwar Economies*. Philadelphia: Temple University Press.

Wolfram, Walt. (1971). Overlapping Influence and Linguistic Assimilation in Second Generation Puerto Ricans. *Paper presented at the American Anthropological Association 1971 Meeting*. New York, New York.

Wolfram, Walt. (1974). *Sociolinguistic Aspects of Assimilation: Puerto Rican English in New York City*. Arlington, VA: Center for Applied Linguistics.

Wood, Simon N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(1):95–114.

Zellou, Georgia, & Tamminga, Meredith. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics* 47:18–35.

Zhao, Sherry. (2010). Stop-like modification of the dental fricative /ð/: An acoustic analysis. *The Journal of the Acoustical Society of America* 128:2009–2020.