Uniformity in phonetic realization:
Evidence from sibilant place of articulation in American English

Eleanor Chodroff[1] and Colin Wilson[2]

[1]University of York, Department of Language and Linguistic Science
Heslington, York YO10 5DD, UK
eleanor.chodroff@york.ac.uk

[2]Johns Hopkins University, Department of Cognitive Science
3400 N. Charles St., Baltimore, MD 21218, USA
colin.wilson@jhu.edu

Uniformity in phonetic realization:
Evidence from sibilant place of articulation in American English

Accepted at Language
Last revised 30 September 2021
Expected publication June 2022

Abstract

Phonetic realization is highly variable and highly structured within and across talkers. We examine three constraints that could structure the phonetic space of related speech sounds: target, contrast, and pattern uniformity. Target uniformity requires a uniform mapping from distinctive features to their corresponding phonetic targets within a talker; contrast uniformity requires a consistent difference in the phonetic targets that realize featural contrasts across talkers; and pattern uniformity requires a uniform template of phonetic targets across talkers. Focusing on American English sibilant fricatives, we measure and compare each constraint's influence on the phonetic targets corresponding to place of articulation. We find that target uniformity is the strongest constraint: each talker realizes a given distinctive feature value in highly similar ways across related sounds. Together with similar findings for other sound classes, this result reveals fine-grained systematicity in the mapping from phonology to phonetics and has implications for theories of speech production and speech perception.*

Keywords: phonetic realization, sibilant fricatives, uniformity, talker variability, Bayesian models

---

*The authors would like to thank Shravan Vasishth for hosting the 2020 Potsdam Summer School on Statistical Methods in Linguistics and Psychology, Lisa Davidson for collecting and sharing the laboratory data, as well as Ryan Cotterell, Matthew Faytak, Josef Fruehwald, and Jane Stuart-Smith for helpful discussion. All data and analyses are available at https://osf.io/bysfa/.

1. Introduction

No one-to-one mapping exists between linguistic units and their phonetic instantiations (Liberman et al. 1967; Massaro 1975; Pisoni and Sawusch 1975). This lack of invariance is a fundamental issue for both the perception and production of language. From the perspective of perception, how do perceivers adapt to extensive variation in the physical signal (whether spoken or signed)? From the perspective of production, how do producers know the limits of acceptable variation for their particular language variety, or even just for intelligibility? It is well-established that variation in phonetic realization is extensive yet structured in many ways (Labov 1972; Miller 1994; Foulkes et al. 2001; Kleinschmidt and Jaeger 2015; Guy and Hinskens 2016; Sonderegger et al. 2020). In the present paper, we explore potential constraints on the mapping from phonological representations, such as segments and their distinctive features, to targets of phonetic realization.

We begin by considering a subinventory of two or more related sounds (e.g. [i] and [u], or [s] and [z]) and their corresponding phonetics targets (i.e. perceptuomotor representations). A given talker could structure the phonetic realization of these sounds by copying a pattern or template of targets that exists in the speech community, adapting it to their anatomy. This scenario allows for talker variation — one speaker's realization of the template may be overall higher or lower on a given phonetic dimension — but otherwise it can be construed as 'maximal phonetic structure'. Provided that the hypothetical template can be adapted for each speaker's anatomy, this system would be fully general across the speaker population. Moreover, clear motivation for such a system exists in speech perception: if each talker has the same template of phonetic targets, perceptual adaptation would involve a simple translation of the pattern for each new talker. This is in fact assumed by many approaches to talker normalization and adaptation, especially for vowel systems (e.g. Lobanov 1971; Nearey 1978; Nearey and Assmann 2007).

In opposition to maximal phonetic structure in the speech inventory, we can consider ‘maximal phonetic bricolage’. Bricolage reflects the constellation of linguistic variables that a talker can exploit for expressing social identity (Eckert 2008; Zimman 2017); taken to the extreme, it would allow talkers to pick and choose phonetic targets independently for each sound. In this scenario, the phonetic space may be structured by overarching social variables, but it would be entirely unstructured within the subinventory and across speakers. For example, the relationship among the phonetic targets of sibilant fricatives like [s], [z], [ʃ], and [ʒ] could be

different for each speaker, depending on how each target is chosen to express some aspect of social identity.

Existing evidence points to an intermediate scenario between these two endpoints, one in which talkers neither copy a single population template nor freely select a target for each individual sound. But what are the constraints on how segments and features are realized phonetically? The present study investigates a set of possible constraints that could account for patterns of structured variation in phonetics, with a focus on the phonetic realization of place of articulation in American English sibilant fricatives.

Structured variation of the type investigated here has been previously observed in various other natural classes. In vowel realization, talker variation is reasonably well-modeled with congruent but shifted vowel templates in the log F1× log F2 space (e.g. Peterson and Barney 1952; Nearey 1978). This suggests highly structured vowel templates across talkers. Systematic differences in stop voice onset time (VOT) are also found within a laryngeal series and among places of articulation across talkers and languages (e.g. Maddieson 1995; Cho and Ladefoged 1999; Chodroff and Wilson 2017; Chodroff et al. 2019). We examine the predictions and strengths of three possible constraints on phonetic realization that could give rise to such between-segment phonetic structure: pattern uniformity, target uniformity, and contrast uniformity. In examining these, we extend previous research on structured variation among speech sounds to the phonetic realization of sibilant place, and investigate whether and how the realization of one sibilant (e.g. [s]) may be systematically linked to that of other sibilants in the subinventory (e.g. [z], [ʃ], and [ʒ]).

1.1. Phonetics–phonology framework

The precise characterization of the phonetics–phonology interface has been a topic of considerable debate (e.g. Nearey 1978; Keating 1988; Browman and Goldstein 1989; Ohala 1990; Pierrehumbert 1990; Kingston 2007; Hamann 2010; Cohn and Huffman 2014; Ladd 2014). The discussion in this paper assumes the following minimal, though not entirely uncontroversial, representational framework that at the core relates discrete phonological units to parametric phonetic representations.

We assume that a phonological surface form is mapped to an abstract set of phonetic targets via a process of phonetic realization. The phonological surface form minimally contains a

sequence of surface phonological segments together with prosodic structure. Each segment is represented by a set of discrete phonological primitives, which are mapped to a corresponding set of continuous phonetic targets. We do not make a strong commitment to the exact nature of the phonological representation, but the primitives necessarily represent phonological contrast and natural class structure among segments. For the sake of clarity, we use classic distinctive features as the phonological primitives for denoting contrast. The phonetic targets are the idealized abstract planning code for the physical instantiation and may be articulatory and/or auditory in form. Critically, each segment has its own constellation of phonetic targets. Finally, the phonetics targets are instantiated via articulation, in spoken language ultimately producing an acoustic signal. Important to note is that the idealized phonetic targets are not the same as any articulatory or acoustic measurements; deviations from the ideal targets arise from random fluctuations and errors. Similar characterizations of the phonetics–phonology interface have also been assumed in Keating (1990), Cohn (1993), Zsiga (1997), Fruehwald (2017), and Volenec and Reiss (2017).

FIGURE 1. Phonetics–phonology interface. Dotted lines indicate abstract representations.

The American English sibilants form a phonologically symmetrical set that contrast in place of articulation and voicing. For simplicity, we employ the [anterior] feature to reflect the contrast between [s z] and [ʃ ʒ], and the [voice] feature to reflect the contrast between [s ʃ] and [z ʒ] (Chomsky and Halle 1968: p. 177; Halle 1983, 1992; Clements 1985). We have chosen the [anterior] phonological feature as the traditional feature that represents the phonological place of articulation contrast, though the 'anterior' label could easily be reFplaced with another label ([distributed], [X], etc.) that classifies [s] together with [z], and separately [ʃ] with [ʒ]. Critically, the phonological representation is underspecified for phonetic information, which must be provided by the process of realization (e.g. Keating 1985; Guenther 1995; Flemming 2004; Volenec and Reiss 2017).[1]

[1] One might wonder whether these are truly phonological features we are investigating or rather some phonetic feature. The present analysis indeed focuses on the feature in terms of how it is

Evidence from cross-linguistic and cross-dialect phonetics indicates that while phonological features may suggest the types of phonetic target employed for a given segment, the features do not fully determine the realization of those targets. Any given feature value says little of the precise location, articulation, or acoustic form of the corresponding phonetic targets: this is instead determined by the phonetic system, which can vary across languages and dialects. Furthermore, as discussed below, speakers have some degree of choice in the targets for the purpose of conveying social or idiosyncratic factors, which could theoretically give rise to wholly independent realizations of [s], [z], [ʃ] and [ʒ]. In contrast, distinctive features could reasonably constrain the specification of phonetic targets, but how and to what degree?

1.2. Variation in the phonetic realization of sibilant fricatives

Variation in the phonetic realization of sibilant fricatives can be observed across a range of contextual factors, as well as from language-, sociolect-, and talker-specific influences. For example, the spectral shape of a fricative is influenced by neighboring vowels, neighboring consonants (Niebuhr et al. 2011), syllable position (Silbert and de Jong 2008), and possibly also speech style (Silbert and de Jong 2008; Maniwa et al. 2009).[2] The extent of coarticulatory influence on the realization of a given segment can vary by language and talker (Solé 1992; Beddor et al. 2002; Yu 2019). The phonetic realization of a given sibilant category also varies considerably across languages (e.g. Nartey 1982; Evers et al. 1998; Gordon et al. 2002; Fuchs and Toda 2010), sociolects (e.g. Flipsen et al. 1999; Stuart-Smith et al. 2003), and talkers (e.g. Haley et al. 2000; Newman et al. 2001; Haley et al. 2010), suggesting some degree of speaker choice in its realization.

Indeed, cross-talker variation in sibilant realization in part reflects anatomical and physiological differences, including the shape and size of the palate, tongue, teeth, lung capacity,

---

mapped to substantive properties; however, this particular feature can also be defined by its distribution in English phonology. Both the place and voice features described above create lexical contrasts and participate in morphophonological alternations, as in the derivational suffix '-ion' (e.g. /s/ ~ /ʃ/: progress ~ progression, compress ~ compression; /z/ ~ /ʒ/: fuse ~ fusion, televise ~ television) or the voicing assimilation in the plural or 3rd person present tense suffixation (e.g. cat[s] ~ dog[z]).

[2] Focused and clear speech conditions numerically raise fricative center of gravity (COG) and less energy is reported below 500 Hz, though the effects among sibilant fricatives may be weaker relative to non-sibilant fricatives.

and airflow regulation. However, talker variation is not wholly reducible to such anatomical and physiological differences. For example, a significantly lower mean spectral center of gravity (COG) was observed for Canadian English [s] (Heffernan 2004) and American English [s] (Li et al. 2007) relative to Japanese [s], and additional differences were observed in the dynamic trajectory of [s] COG between English and Japanese (Reidy 2016). Fuchs and Toda (2010) also identified significant differences between English and German [s] in its acoustic and articulatory instantiation. While the physical morphology of speakers could differ in minor ways across language populations (Fuchs and Toda 2010; Dediu et al. 2019), such acoustic differences more likely reflect cross-linguistic variation in the phonetic targets for sibilant place (e.g. the constriction location, degree of constriction, and these target dynamics) because they are larger than would be expected due to anatomical differences alone (Gordon et al. 2002; Fuchs and Toda 2010).

Additional evidence for talker-specific control in the phonetic realization of sibilants comes from sociophonetics. Within a language, sibilants vary according to sociolinguistic variables such as gender (American English: Strand and Johnson 1996; Flipsen et al. 1999; Podesva and Van Hofwegen 2014; Canadian English: Heffernan 2004; Glaswegian English: Stuart-Smith et al. 2003; British English: Levon and Holmes-Elliot 2013; mixed Australian, North American, and UK talkers: Fuchs and Toda 2010), sexual orientation (Linville 1998), age (Stuart-Smith et al. 2003; Podesva and Van Hofwegen 2014), socioeconomic class (Stuart-Smith et al. 2003), and region (Podesva and Van Hofwegen 2014). Again, these phonetic differences could arise from population-level differences in speaker anatomy; however, it is more likely that the precise articulation of [s] conveys the talker-specific expression of a sociolinguistic variable.

Moreover, spectral differences in sibilant fricatives that covary with gender extend beyond anatomical explanations. On average, women have shorter vocal tracts than men (Schwartz 1968), but this dimorphism is primarily found posterior to the typical constriction locations for sibilants (Strand 1999). Even after controlling for palate size and length, Fuchs and Toda (2010) found that female speakers had a more fronted articulation of [s] than male speakers in both English and German language groups. Taken all together, these observations indicate some degree of implicit talker choice in the phonetic realization of sibilant fricatives, as for other speech sounds.

1.3. Constraints on phonetic realization

Evidence from cross-linguistic, cross-dialectal, and cross-speaker variation implies a range of permissible phonetic realizations for each segment. Principles of sufficient contrast (e.g. phonetic dispersion or perceptual distinctiveness; Liljencrants and Lindblom 1972; Lindblom 1986; Flemming 2004) and articulatory ease (e.g. Lindblom and Maddieson 1988; Napoli et al. 2014) likely restrict the range in important ways. However, many employable phonetic targets are apparently available even after these constraints have been applied. 'Maximal bricolage' would permit independent phonetic targets for each segment while still satisfying sufficient contrast and ease. This would support the expression of talker identity but leave little systematicity in the realization of related sounds.

Conversely, talkers could maximize consistency in the mapping from phonological segments to phonetic targets ('maximal phonetic structure'). A fully constrained mapping would have talkers reuse a standard template of targets for similar sounds. This tight regulation of phonetic realization on the part of the speaker would have clear benefits for the listener. If each talker within a population maintains the same pattern of phonetic realizations, the listener could simply shift the template up or down to adapt to a given talker; for example, once a listener identifies how e.g. [s] is produced, the realizations of the other sibilant fricatives by the same talker could be read off the template. We call this constraint on phonetic realization PATTERN UNIFORMITY.

> PATTERN UNIFORMITY: across speakers of a language, the difference between phonetic targets for phonological surface segments k_1 and k_2 should be identical

The notion of pattern uniformity is akin to several talker normalization and adaptation algorithms that assume consistent relationships of phonetic variables across speakers (e.g. Joos 1948; Nearey 1978; McMurray and Jongman 2011). For example, Nearey (1978) outlined a constant ratio hypothesis for vowels in which the ratio of F1 and F2 values should be constant across talkers; in log space, this is translated as the constant log-interval hypothesis and expressed as a sliding template of vowel categories in the log F1× log F2 space (e.g. Nearey and Assmann 2007). Pattern uniformity extends this principle beyond vowel formants to apply more generally to the phonetic realization of phonological segments.

As stated, pattern uniformity does not place any restriction on how similar or distinct the phonetic targets of different speech sounds should be. For example, any template of sibilant targets would be permissible provided all talkers conformed to that pattern. While pattern uniformity may play a role in restricting variation among phonetic targets across talkers, it does not restrict the degree of similarity among phonetic targets of segments that share or contrast on distinctive features.

Alternatively, it may be useful to focus on each aspect of a pattern as opposed to the whole. To this end, we consider how the composition of a surface segment, that is, its distinctive features, may directly constrain phonetic realization. We formalize two constraints that could govern the mapping from distinctive feature values to corresponding phonetic targets: TARGET UNIFORMITY and CONTRAST UNIFORMITY.

> TARGET UNIFORMITY: within each speaker of a language, the phonetic targets corresponding to phonological feature value [αF] should be identical for all segments that are specified [αF] (where α can be + or – for binary features)

> CONTRAST UNIFORMITY: across speakers of a language, the differences of phonetic targets corresponding to different values of a feature [F] should be identical

The first constraint, target uniformity, requires that the phonetic targets for segments that share a feature value be identical within a talker. Target uniformity builds on a line of previous and related principles posited in the literature that emphasize reuse of phonetic targets corresponding to phonological primitives (Maddieson 1995; Keating 2003; Ménard et al. 2008; Guy and Hinskens 2016; Chodroff and Wilson 2017; Fruehwald 2017). For instance, gestural economy requires reuse of individual gestures across multiple speech sounds (Lindblom 1983; Lindblom and Maddieson 1988; Maddieson 1995). A similar notion of uniformity has been motivated by the study of allophonic variation (Keating 2003): for example, speakers may alternatively prioritize articulatory or acoustic uniformity in the instantiation of stop voicing, despite potentially increased articulatory difficulty. The Maximal Use of Available Controls (MUAC) principle requires reuse of gestural or perceptual controls in the implementation of a

distinctive feature across segments with that feature (Schwartz et al. 2007; Ménard et al. 2008).[3] For example, Ménard et al. (2008) observed a high degree of talker-specific F1 stability across vowels with a shared height feature. From the perspective of sound change, Fruehwald (2013) also proposed that parallel shifts of phonetic targets over time may arise from a shifted phonetic implementation of a distinctive feature. In other words, the shifting phonetic targets may be yoked to a single distinctive feature shared across segments (see also Fruehwald 2017).

We have documented strong covariation of talker mean VOT among aspirated stop consonants across speakers of American English (Chodroff and Wilson 2017, 2018), as well as among stops with a shared laryngeal feature across over 100 typologically diverse languages (Chodroff et al. 2019). The observed covariation in VOT is highly indicative of structure in the underlying phonetic targets. Such covariation could plausibly arise from underlying identity in the phonetic realization of the shared laryngeal feature; assuming a consistent glottal gesture with a consistent timing relationship to the oral release, minor differences in VOT among place of articulation can be accounted for by biomechanical factors (Löfqvist and Yoshioka 1984).

Each of these proposals shares the intuition that each phonological primitive should have a uniform phonetic realization. Thus far, this principle has been treated categorically: phonetic targets (or gestures) should be identical across segments specified with the relevant phonological primitive. In its purest form, target uniformity matches many of these previous proposals, but in contrast to previous accounts, we reposition all of the constraints considered here as violable influences on phonetic realization. Between segments, major deviations of phonetic targets of a shared feature are highly improbable, whereas minor deviations are acceptable. One of the principal goals in later sections of the paper is to identify just how strongly each constraint influences phonetic realization.

The second uniformity constraint under consideration regulates segments that differ in the specification of a phonological feature. Contrast uniformity resembles pattern uniformity in its utility for listener adaptation, but focuses instead on the phonetic realization of phonological

---

[3] The Maximal Use of Available Controls is a development from the phonological principle of the Maximal Use of Available Features (MUAF; Ohala 1979, 1980), which is also related to the notion of feature economy (Clements 2003). Similar to the above cited proposals, we distinguish the phonetic from the phonological level, and the particular constraints that apply to each. Overall, the selection of features within a language inventory says little about the phonetic realization by a given talker.

contrast. Given the centrality of contrast in linguistic systems, we find this a relevant influence to explore. However, contrast uniformity is unlikely to hold in its strictest form: for example, previous studies have demonstrated significant cross-speaker variability in the degree of contrast between speech sounds, and even between the constriction locations for [s] and [ʃ] (e.g. Newman et al. 2001; Ghosh et al. 2010; Yunusova et al. 2012; see also Chodroff and Wilson 2017 for stop consonant VOT). Given such findings, many instead have argued for a principle of sufficient phonetic contrast, as opposed to maximal phonetic dispersion (e.g. Lindblom 1986): contrast uniformity simply requires that talkers replicate whichever point on this continuum is used in the population. We retain contrast uniformity in our evaluation, but expect this constraint to be more tolerant of violation than target uniformity.

We adopt the strong position that constraints on phonetic realization, whatever precise form they take, should be universal and apply to all feature–target pairings. However, as already noted and unlike some previous proposals, we do not necessarily expect categorical restrictions on phonetic realization. Talkers may violate a uniformity constraint in the realization of segment due to intrasegmental coarticulation, pressure from competing constraints such as perceptual distinctiveness, or the use of phonetic variables for social expressivity. Intrasegmental coarticulation refers to a strong influence of and interaction between multiple distinctive features in the phonetic realization of a segment (Volenec and Reiss 2017). Our expectation is that, particularly for target uniformity, these violations should be minimal.

In the present study, we focus on the phonetic realization of place of articulation across sibilant fricatives in two datasets of American English: isolated speech recorded in a laboratory environment and spontaneous speech from the Buckeye Corpus (Pitt et al. 2005). We investigate the strengths of target, contrast, and pattern uniformity on the selection of phonetic targets within a Bayesian framework using prior sensitivity analysis (Vanpaemel 2010; Kary et al. 2016). This type of analysis determines the influence of prior specification on posterior estimation, and can be used to select among competing, quantitatively specified models. As described in the following section, we model phonetic realization using a Bayesian linear mixed-effects regression model of a key phonetic correlate of sibilant place of articulation, and model the uniformity constraints as prior distributions of relevant population and talker-specific parameters. The strength of each uniformity constraint can then be assessed using Bayes factors, which provide an index of the relative evidence in favor of one model over another.

2. Evaluating uniformity

In the following sections, we first explain our acoustic-phonetic measure of the phonetic target for sibilant place of articulation, the spectral mid-frequency peak. We then describe how the constraints of target, contrast, and pattern uniformity are formalized in regression analyses. Specifically, we introduce a Bayesian linear mixed-effects model of phonetic realization (see also Vasishth et al. 2018) that allows us to assess the strength of each constraint.

2.1.1. Phonetic correlate to sibilant place of articulation

The uniformity constraints are assumed to operate on the mapping from distinctive features (e.g. values of [anterior]) to phonetic targets (e.g. the phonetic realization of sibilant place). As the phonetic target cannot be measured directly, a phonetic correlate of the target must instead be selected. While the [anterior] feature of fricatives could have a complex set of phonetic targets, a principal one is the articulatory location of the constriction. The present analysis thus employs an acoustic-phonetic correlate of the constriction location. Though a direct, articulatory measure of the constriction location could be advantageous, it would still only approximate the underlying phonetic target for place of articulation. Moreover, acoustic analysis is quite scalable, allowing for large-scale investigation of the question.

Several acoustic correlates of place of articulation have been employed in the literature. For example, COG and spectral peak have been widely used as correlates of fricative place; however, they do not cleanly separate components of the spectrum that arise separately from the source and filter, even after using mitigating techniques like high-pass filtering (see Koenig et al. 2013). We instead adopt the spectral mid-frequency peak as the best available phonetic correlate to sibilant place of articulation with the general caveat that no phonetic measure, acoustic or articulatory, can perfectly reveal a phonetic target. The mid-frequency peak has been proposed as an alternative and more precise acoustic measure of fricative place relative to common measures such as COG or spectral peak (Koenig et al. 2013; Shadle et al. 2014). The measure reflects the resonance of the vocal tract cavity anterior to the constriction, and therefore approximates the location of tongue constriction (Shadle et al. 2016). It is also known to be relatively unaffected by source properties such as vocal fold vibration and vocal effort (Koenig et al. 2013). In previous studies, mid-frequency peak was defined as the peak frequency between 3000 and 7000

Hz for the alveolar sibilants; however, that interval was defined based on visual inspection of the lowest peak frequency in [s] based on the study sample of adolescent speakers. The mid-frequency peak has also not previously been defined for postalveolar sibilants which have a larger anterior cavity and a correspondingly lower resonant frequency.

Based on visual inspection of the data from the laboratory study reported below, we identified an estimate of the mid-frequency peak that closely corresponded to the lowest salient spectral peak above any voicing excitation.[4] Following previous literature on sibilant analysis, a multitaper spectral analysis was conducted over the middle 20 ms of each sibilant in prevocalic position (Shadle and Mair 1996; Reidy 2015, 2016). Within that spectrum, the mid-frequency peak was defined as the frequency of maximum amplitude between 2000 and 6000 Hz if the corresponding power spectral density exceeded 1 $\mu Pa^2/Hz$, and otherwise, as the frequency of maximum amplitude between 3000 and 7500 Hz.[5] Sizable peaks below ~6000 Hz were frequently accompanied by a secondary peak above 6000 Hz; the mid-frequency peak would be the first of these two. This frequency was then converted from hertz to the psychoacoustic scale of equivalent rectangular bandwidth (ERB) for closer approximation of the perceptual representation (Glasberg and Moore 1990).

One notable benefit of the mid-frequency peak, especially in comparison to COG or spectral peak, is its applicability at lower sampling rates: the measure explicitly ignores any high-frequency excitations that are commonly present in sibilant fricatives. As sibilants can contain substantial high frequency energy, sampling rate can affect measures such as COG (Shadle and Mair 1996), while our primary measure of mid-frequency peak, confined to fall within the 2000 to 7500 Hz frequency range across all sibilants examined here, is effectively invariant across sampling rates at or above 16 kHz. For comparison with previous studies and for interpretability, we also report descriptive statistics and correlations of the mid-frequency peak and center of gravity (COG) after high-pass filtering at 550 Hz (Forrest et al. 1988; Koenig et al.

---

[4] Visualization of the spectra can be further investigated by setting up the Shiny app available for download here: https://osf.io/bysfa/.

[5] Multitaper spectral analysis is an alternative to the more conventional Fourier analysis for spectral density estimation that uses multiple tapers to provide independent estimates of spectral information (Thomson 1982; Blacklock 2004). The tapers are windows over the signal, and in the particular analysis are Slepian and orthogonal. Unlike Fourier analysis, the technique does not make a strong assumption of periodicity in the signal. The analysis had 8 tapers and a time bandwidth of 4.0 and was implemented using the multitaper R package (Rahim and Burr 2020).

2013; see Appendix). This high-pass filter was only relevant for the COG measure and was used to minimize the influence of voicing as much as possible. The COG estimates from the two corpora should not be compared directly against each other as their sampling rates differ.

2.1.2. Computational analysis

Quantitative evaluation of the uniformity constraints involves a first-pass assessment of the degree to which talker-specific mid-frequency peak means are correlated among sibilant fricatives across speakers. Target uniformity predicts strong correlations between segments with a shared [anterior] specification (e.g. [s] and [z]; [ʃ] and [ʒ]) that arise from underlying identity. Contrast uniformity predicts strong correlations between segments contrasting in the [anterior] feature (e.g. [s] and [ʃ], as well as [z] and [ʒ]); the strength of these correlations would arise from a consistent difference in phonetic realization across speakers. If strong correlations are observed among all four sibilants, pattern uniformity is automatically achieved.

We employ a Bayesian linear mixed-effects regression to model the variation in sibilant mid-frequency peak that arises in the phonetic realization of phonological surface segments, as well as variation from contextual, social, and talker influences (Vasishth et al. 2018). In Bayesian inference, the aim is to identify a posterior distribution, that is, the probability of the model parameters given the observed data. The posterior is approximated from the likelihood of the data (the probability density of the observed data given the model parameters) multiplied by the prior probability density of the parameters.[6]

In addition to the effects of phonological features, acoustic-phonetic measures will naturally be influenced by the phonetic context, gender, and idiosyncratic features of the talker. In predicting mid-frequency peak (ERB), we use the following independent variables:

[6] The brms package in R was used for all model fitting and comparison (Bürkner 2017, 2018). This package provides an R interface to the Stan programming language, which uses the No-U-Turn Sampler for parameter estimation (Hoffman and Gelman 2014). Each model was run for 50,000 iterations: the first half of the samples were discarded as burn-in, and the second half formed the posterior distribution.

DISTINCTIVE FEATURES. Fixed effects of [anterior], [voice], and the interaction between [anterior] and [voice][7]

CONTEXTUAL FEATURES. Fixed effects of following vowel height, following backness, and the interaction between height and backness

SOCIAL FEATURES. Fixed effect of talker gender

TALKER FEATURES. Random intercept for talker, random slopes for place, voice, and the interaction between place and voice

Together, these form the following model structure in (1).

$$(1)\ y_{i,j} \sim \beta_0 + \beta_{anterior}x_{anterior,i} + \beta_{voice}x_{voice,i} + \beta_{anterior:voice}x_{anterior:voice,i} + \beta_{height}x_{height,i} + \beta_{front}x_{front,i} + \beta_{height:front}x_{height:front,i} + \beta_{gender}x_{gender,i} + \mu_{0,j} + \mu_{anterior,i,j} + \mu_{anterior,i,j} + \mu_{place:voice,i,j} + \varepsilon_{i,j}$$

β reflects parameter estimates of fixed effects, μ reflects parameter estimates of random effects, ε reflects the error term, *i* corresponds to individual data points, *j* corresponds to individual talkers, and the colon to an interaction. In the present analysis, all predictors are categorical with two levels that are weighted effect coded, in which one level is assigned a weight of +1 and the other is determined based on the relative sample size. The exact coding for each parameter and corpus is reported in the Appendix.

The second analysis assesses the approximate strength of each uniformity constraint while accounting for overall variation in the data. Using a Bayesian approach, we can increase the strength of a uniformity constraint by modifying the breadth of the relevant prior probability distribution. The prior should place some constraint on the selection of phonetic targets, even if the speaker or population ultimately deviates from this instantiation. Regardless, some priors will be more consistent with the data than others. For example, target uniformity as applied to the feature [anterior] predicts minimal influence of [voice] specifications on the phonetic realization of place of articulation. The prior distribution of [voice] could then be modeled as a normal distribution, centered on 0, indicating no difference in mid-frequency peak between e.g. [s] and [z], and with a very small standard deviation, reflecting a low tolerance for any violation of

[7] We use the terms place and voice without brackets to refer to the model factors corresponding to the features [anterior] and [voice].

target uniformity. However, if target uniformity plays virtually no role in target specification, then a uniform prior distribution of [voice] should be more compatible with the data. (The uniform distribution places equal probability over a range of differences in the phonetic targets for constriction location between [s] and [z].) Previous work implementing a similar approach via comparison of prior distributions is described in Vanpaemel (2010) and Kary et al. (2016).

To assess the strength of uniformity constraints, we directly relate each one to a component in the linear mixed-effects model and modulate the strength of its prior. We then use Bayes factors to compare models that differ in the prior distributions. The evaluation of target, contrast, and pattern uniformity is implemented in four parts: the first two sets of comparisons investigate target uniformity, the third set contrast uniformity, and the fourth set pattern uniformity. In this setting, target uniformity has scope over the population-level effect of [voice] and the random by-talker slope for [voice]. Contrast and pattern uniformity correspond to talker-specific shifts with respect to the estimated population means.

TARGET UNIFORMITY: POPULATION. In the first set of models, we examine the strength of target uniformity on the population-level effect of [voice]. (The predictions of contrast and pattern uniformity do not involve population-level effects, as they deal solely with cross-talker differences.) As described above, target uniformity predicts minimal influence of [voice] on the phonetic realization of the [anterior] feature, here measured by the mid-frequency peak. We formalize this restriction in terms of the prior distribution of the [voice] factor. Specifically, we test five prior distributions on [voice]: four normal distributions with mean of 0 ERB and standard deviations {0.01, 0.1, 0.5, 1} ERB, and a uniform prior distribution with equal probability over the real numbers from -10 to +10 ERB. In subsequent sections, we will refer to the prior probability distributions using their shorthand forms: for the normal distribution, this is $\mathcal{N}(mean, standard\ deviation)$; for the uniform distribution, this is $Unif(minimum, maximum)$. The priors over the random intercepts and slopes for talker are implemented as normal distributions, centered on 0 ERB with a standard deviation of 1 ERB, corresponding to the largest normal standard deviation considered in our model comparisons.

TARGET UNIFORMITY: TALKER. In the second set of models, we manipulate the prior distribution of the random slope for [voice] while holding all other priors constant. This further tests the influence of target uniformity on phonetic realization: the influence of [voice] should be

minimal for any particular speaker, just as it should be at the population level. We test four prior distributions of the by-talker slope for [voice]: $\mathcal{N}(0, 0.01), \mathcal{N}(0, 0.1), \mathcal{N}(0, 0.5), \mathcal{N}(0, 1)$.

CONTRAST UNIFORMITY: TALKER. Contrast uniformity stipulates that the effect of the [anterior] factor should be the same across speakers. While we expect a clear contrast in mid-frequency peak between [+anterior] and [-anterior] sibilants in the population, individual speakers should not deviate from this effect. As such, the random by-talker slope for [anterior] should not vary across speakers. In this analysis, we compare models that differ in the prior distribution over the random by-talker slope for [anterior]. We test four prior distributions of the by-talker slope for [anterior]: $\mathcal{N}(0, 0.01), \mathcal{N}(0, 0.1), \mathcal{N}(0, 0.5), \mathcal{N}(0, 1)$.

PATTERN UNIFORMITY: TALKER. Pattern uniformity stipulates that talkers should not stray from the population template, with the critical exception that their targets can be translated, in lockstep, on the relevant phonetic dimensions. For this analysis, we modulate the prior distributions of the random by-talker slopes for [anterior], [voice], and the interaction between [anterior] and [voice]. In essence, the only way speakers should differ from each other is in the absolute value—the intercept—and speaker-specific influences of [anterior] and [voice] should be minimal to nonexistent. We test four sets of prior distributions of the by-talker slopes for [anterior], [voice], and the interaction between [anterior] and [voice]: $\mathcal{N}(0, 0.01), \mathcal{N}(0, 0.1), \mathcal{N}(0, 0.5), \mathcal{N}(0, 1)$.

For all other independent variables, we used weakly informative prior distributions based on previously reported sibilant measures from the Jongman (2000) American English fricative dataset (TABLE 1). The dataset contained phonetic measures of fricatives from 20 native speakers of American English, producing fricative-initial CVC syllables in a laboratory setting; recordings were sampled at 22 kHz. As our main effects were each two-level categorical predictors that were weighted effect coded, the prior distributions represent the difference between the first-listed level (assigned a weight of +1) and the sample mean. In the highly balanced Jongman (2000) dataset, we approximated this as half the difference between contrasting means. For example, we estimated a prior distribution of $\mathcal{N}(2, 5)$ for the effect of place of articulation on mid-frequency peak in ERB. Jongman et al. (2000) found a spectral peak difference of 5.3 ERB (2979 Hz) and a COG difference of 3.4 ERB (2186 Hz) between [+anterior] and [-anterior] sibilants. Because the sample sizes for [+anterior] and [-anterior] sibilants were balanced in that study, we then take half of the estimated difference as the model estimate for place of

articulation (i.e. 2.65 ERB for spectral peak and 1.7 ERB for COG). Given the mismatch in measure between our study and theirs, we added an additional approximation, rounding to an estimated mean $\hat{\beta}_{anterior}$ of 2 ERB. A similar procedure was implemented for each additional effect and interaction. The priors were normal distributions with means marginally shifted from 0, if at all, and broad standard deviations that minimized the consideration of unexpected or physically impossible estimates of a contrast.

TABLE 1. Prior distributions on model components that remain fixed during comparisons. The main effects are each categorical predictors with two levels that have been weighted effect coded according to their sample size. This coding scheme reflects the difference between the first-listed level (assigned a weight of +1) and the sample mean. The mean specifications are loosely based on previously reported spectral peak and COG measures from Jongman et al. (2000) that have been converted to ERB. In all cases except the intercept, the standard deviations were broad at 5 ERB, but still informative for narrowing the range of values considered by the model. For reference, the difference between 7000 Hz and 4000 Hz is equivalent to 5 ERB. The ERB differences become smaller as the input values move up on the hertz scale, and greater as the input values move down on the hertz scale.

Predictor	**Prior** ERB	**Levels**	**Mean Spectral Peak** Hz	ERB	**Mean COG** Hz	ERB
Intercept	$\mathcal{N}(0, 1)$	—	—	—	—	—
[anterior]	$\mathcal{N}(2, 5)$	[+anterior] [–anterior]	6809 3830	31.6 26.3	6817 4631	31.7 28.3
[anterior] x [voice]	$\mathcal{N}(0, 5)$	[s, ʒ] [z, ʃ]	5316 5324	28.9 28.9	5678 5769	29.9 30.1
following vowel height	$\mathcal{N}(0, 5)$	[+high] [–high]	5320 5319	28.9 28.9	5713 5729	30.0 30.0
following vowel backness	$\mathcal{N}(0.15, 5)$	[+front] [–front]	5445 5193	29.2 28.7	5811 5636	30.1 29.9
vowel height x backness	$\mathcal{N}(0, 5)$	[+fr, +hi], [–fr, –hi] [+fr, –hi], [–fr, +hi]	5283 5356	28.9 29.0	5690 5758	29.9 30.0
gender	$\mathcal{N}(1, 5)$	female male	5895 4744	30.0 27.9	6241 5206	30.8 29.1

3. Uniformity in isolated laboratory speech

The predictions of target, contrast, and pattern uniformity were first tested in a laboratory corpus of fricative-initial syllable productions from 22 native speakers of American English. The corpus contained an approximately equal number of tokens for each sibilant fricative ([s z ʃ ʒ]) in matched segmental contexts. The high degree of control in the stimuli contributed to the goal of isolating the potential sources of variation and covariation due primarily to talker differences.

3.1. Methods

3.1.1. Participants

Twenty-two participants (15 female) were recruited at New York University for an experiment on non-native consonant cluster production and perception, in which one of the tasks was a fricative-initial syllable production task. All participants were native speakers of American English. Participants were given monetary compensation for participation.

3.1.2. Materials and procedure

Participants recorded fricative-initial consonant-vowel-consonant (CVC) syllables in isolation during an unrelated experiment on the perception and production of non-native consonant clusters. All recordings were made with a Zoom H4n digital recorder and an Audio-Technica ATM-75 head-mounted condenser microphone in a sound-attenuated booth at a sampling rate of 44.1 kHz. The CVC syllables were composed by fully crossing the fricatives [ð θ f v s ʃ z ʒ] with the vowels [i ɪ eɪ ɛ æ a ɔ oʊ ʊ u ʌ], and [t] (Jongman et al. 2000). Two [ʃ]-initial combinations were excluded due to their profane nature. Only syllables beginning with the sibilant fricatives [s ʃ z ʒ] were considered for analysis. In some cases, participants could not interpret the orthographic mapping for [ʒ] and [ð]: two participants (1 female, 1 male) did not produce any instances of [ʒ].

Each trial in the experiment consisted of three parts. Participants heard a prerecorded multisyllabic nonword with an initial consonant cluster for an unrelated experiment, and then produced a CVC syllable as a distractor item, followed by a reproduction of the auditorily presented nonword. The distractor items were the fricative-initial CVC syllables analyzed here. Each was presented visually on a monitor with a standard grapheme-to-phoneme mapping. There were 12 unique presentation orders, and each CVC syllable was presented two to three times.

116 mispronounced tokens were excluded. A total of 1,926 sibilants remained for analysis, with the median per talker and sibilant ranging from 21 to 24 (Appendix, TABLE 6).

3.1.3. Data preparation

Phonetic segmentation was performed with the Penn Phonetics Lab Forced Aligner (Yuan and Liberman 2008). All boundaries were manually corrected to align to the fricative onset and offset, which were defined by the presence of frication. This often coincided with the onset of periodicity in the vowel, but in cases when periodicity and frication overlapped, the boundary was placed after frication ended.

3.2. Results

The sibilant-specific grand means for mid-frequency peak largely reflected the similarity and contrast in anteriority: sibilants with a shared place of articulation specification had comparable mid-frequency peak means, and as expected, sibilants that contrasted in place of articulation differed substantially from one another (Appendix, TABLE 7 for talker means and standard deviations of mid-frequency peak in ERB, mid-frequency peak in Hz, and COG in Hz). The variation across talker mean mid-frequency peaks and standard deviations was sizable for each sibilant. These parameters were weakly to moderately correlated with each other in ERB, but did not reach significance (see Appendix, TABLE 8 for correlations between talker means and standard deviations within each sibilant category). The correlations between the talker-specific mean and standard deviation contrast in strength and directionality with those observed in many temporal measures, which are generally moderate to strong, and positive (e.g. Byrd and Saltzman 1998; Shaw et al. 2009; Turk and Shattuck-Hufnagel 2014; Chodroff and Wilson 2017).[8]

3.2.1. Correlation analysis

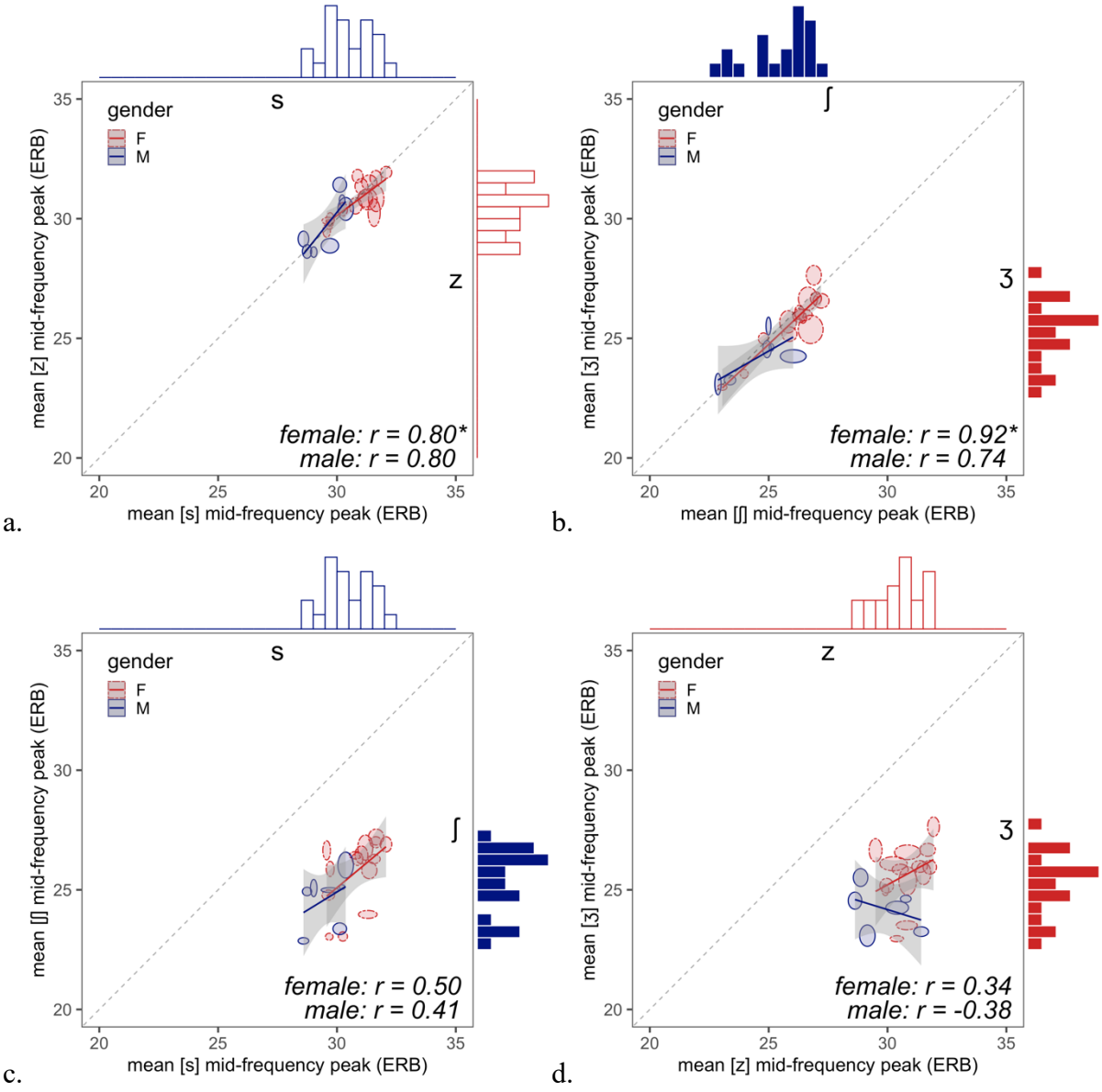
As a first analysis of the influence of the uniformity constraints, we examine the general patterns of covariation of talker mid-frequency peak means among the sibilant categories. Strong

[8] Correlations are described using modifiers based on recommendations in Evans (1996): a 'weak' correlation describes a coefficient below 0.40, a 'moderate' correlation means the coefficient is between 0.40 and 0.59, and a 'strong' correlation means the coefficient is above 0.59.

correlations among all categories would support the pattern uniformity constraint. Strong correlations between sibilants with the same place of articulation would be consistent with the predictions of target uniformity, but a correlation alone does not necessarily entail underlying identity. This latter part is investigated in the Bayesian analysis. Finally, strong correlations between sibilants contrasting in place of articulation would be consistent with an influence of contrast uniformity.

Because of the bimodality in talker-specific mid-frequency peak means across male and female speakers, correlations were calculated separately for each gender. As shown in FIGURE 2, talker-specific mid-frequency peak means were strongly correlated between homorganic sibilants within each gender group ([s] – [z] female: $r = 0.80$, male: $r = 0.80$; [ʃ] – [ʒ] female: $r = 0.92$, male: $r = 0.74$; see Appendix, TABLE 9 for correlations of mid-frequency peak in ERB, mid-frequency peak in Hz, and COG in Hz). While consistently strong, only the correlations of female speaker means reached significance (each $p < 0.001$); the strong but non-significant correlations of male speaker means were likely attributable to the low sample size of seven speakers. Correlations between sibilants contrasting in place were considerably weaker than those between homorganic sibilants, and were not significantly different from zero ([s] – [ʃ] female: $r = 0.50$, male: $r = 0.41$; [z] – [ʒ] female: $r = 0.34$, male: $r = -0.38$; each $p > 0.001$). Given the strong correlations among homorganic sibilants, this pattern of results is consistent with a sizable influence of target uniformity, but less so for contrast or pattern uniformity.

FIGURE 2. Variation and covariation of sibilant mid-frequency peak (ERB) across talkers in the American English isolated speech data. Each ellipsoid is centered on a pair of talker-specific means and is color-coded by talker gender; the size of the ellipsoid reflects 1/5 of the standard deviation of the respective sibilants. Marginal histograms indicate the variation in talker means for each sibilant category. The asterisk indicates $p < 0.01$. Gray shading reflects the local confidence interval around the best-fit linear regression of talker means for each gender.



3.2.2. Bayesian analysis

In the Bayesian analysis, we investigate the presence and strength of each uniformity constraint on the phonetic realization of sibilant fricatives in the laboratory speech data. As outlined in Section 2, we compare a series of models with differing prior distributions over the parameters relevant for uniformity using Bayes factors. These are interpreted using Jeffreys' scale in which a factor between 1 and 3 reveals 'anecdotal' evidence for M_1, a factor between 3 and 10 reveals 'moderate' or 'substantial' evidence for M_1, a factor between 10 and 30 reveals 'strong' evidence for H_1, a factor between 30 and 100 reveals 'very strong' evidence for H_1, and a factor over 100 reveals 'extreme' or 'decisive' evidence for M_1 (Jeffreys 1961; Nicenboim et al. 2021).

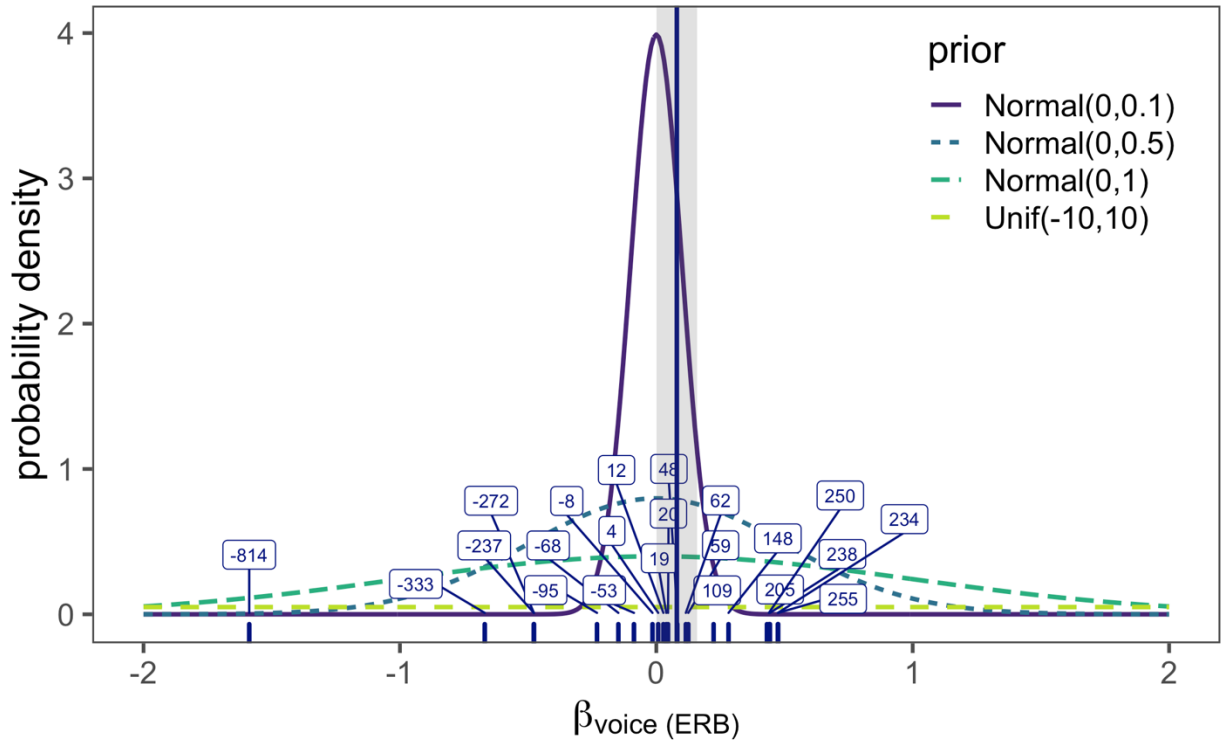
3.2.2.1. Target uniformity

The first set of model comparisons investigates target uniformity, manipulating the prior distribution of the fixed effect of [voice]. A strong target uniformity constraint would minimize the influence of [voice] on the phonetic realization of each sibilant fricative, as the [anterior] should have dominant control over the place of articulation target. This would correspond to a prior distribution of [voice] centered on zero with a small standard deviation. As described in Section 2, a range of prior distributions was tested: four normal distributions centered on 0 with varying standard deviations, and a uniform distribution over values from –10 to +10 ERB: $\mathcal{N}(0, 0.01), \mathcal{N}(0, 0.1), \mathcal{N}(0, 0.5), \mathcal{N}(0, 1)$, and $Unif(-10,10)$. The uniform distribution places equal probability over all possible estimates of [voice] within that range. A depiction of these prior distributions on the [voice] contrast is shown in FIGURE 3 along with the mean mid-frequency peak deviation of the [-voice] sibilants from the talker-specific mean.

As shown in TABLE 2a, strong to very strong evidence exists in favor of models with normally distributed priors over [voice] with standard deviations less than or equal to 1, relative to the model with a uniform prior over [voice]. Of the normally distributed priors centered on zero, those with smaller standard deviations are generally preferred. In particular, moderate evidence exists in favor of the model with a prior of $\mathcal{N}(0, 0.1)$ over [voice] relative to comparable priors with larger standard deviations. Very little difference is found between models with priors of $\mathcal{N}(0, 0.01)$ and $\mathcal{N}(0, 0.1)$, but with anecdotal evidence towards the broader prior of $\mathcal{N}(0, 0.1)$.

The second set of model comparisons also investigates the role of target uniformity, but with respect to the random by-talker slope for [voice]. The previous set of comparisons indicated that at the population level, the data are more consistent with models that have priors favoring a very minimal influence of [voice] relative to models that allow for more variation in that effect. However, wide variation across speakers could exist in exactly how strongly they conform to this constraint. As shown in TABLE 2b, talkers indeed vary in just how strongly they conform to this population norm. For the random by-talker slope for [voice], anecdotal to moderate evidence is found in favor of models with priors of $\mathcal{N}(0, 0.1)$ and $\mathcal{N}(0, 0.5)$ relative to models with a stronger prior, $\mathcal{N}(0, 0.01)$, or a weaker prior, $\mathcal{N}(0, 1)$. Between these two models, moderate evidence is found in favor of $\mathcal{N}(0, 0.1)$ relative to $\mathcal{N}(0, 0.5)$.

FIGURE 3. Priors over the population effect of [voice]. Given the coding scheme, the prior reflects the distance of the [-voice] mid-frequency peak from the mean. (The effect of [voice] was weighted effect coded to standardize the procedure across corpora: [-voice] = +1, [+voice] = -1.04.) The tightest prior of $\mathcal{N}(0, 0.1)$ is not pictured here due to its concentrated probability density. The rug plot corresponds to half the difference between by-talker [-voice] and [+voice] mid-frequency peak means in ERB for the American English isolated speech data. These are labeled with their corresponding contrast in hertz. (Empirical by-talker differences range from -1628 Hz to 510 Hz.) The vertical line reflects the estimated mean effect of [voice], 0.08, using the model reported in Section 3.2.2.4. The gray shading represents the 95% credible interval around that estimate ([0.00, 0.16]).



3.2.2.2. Contrast uniformity

The third set of model comparisons investigates the strength of contrast uniformity by modulating the prior distribution over the random by-talker slope for [anterior]. The population-level contrast between heterorganic sibilants is specified in the fixed effect of [anterior]; contrast uniformity stipulates that talkers should not deviate from that population effect. By modifying

the prior distribution over the random by-talker slope for [anterior], we can investigate whether the data is more consistent with models that tightly constrain this variation or models that allow greater freedom in the contrast. As shown in TABLE 2c, decisive evidence is found in favor of models with larger standard deviations ($\mathcal{N}(0, 0.5)$, $\mathcal{N}(0, 1)$) relative to comparable models with smaller standard deviations for that prior distribution. Anecdotal evidence is, however, found in favor of a prior of $\mathcal{N}(0, 0.5)$. over the by-talker slope for [anterior] relative to a prior of $\mathcal{N}(0, 1)$, suggesting a potential upper limit on cross-talker variation in the effect of [anterior] on mid-frequency peak.

3.2.2.3. Pattern uniformity

The fourth and final set of model comparisons investigates the strength of pattern uniformity, or overall consistency in the implementation of the population-level template for mid-frequency peak across talkers. To investigate this, the standard deviations of the prior distributions over the random by-talker slopes for [anterior], [voice] and [anterior] × [voice] are modulated while the prior distribution over the random by-talker intercept is kept relatively large at $\mathcal{N}(0, 1)$. As shown in TABLE 2d, decisive evidence is found in favor of models having wider standard deviations in the prior distributions over these random slopes, except at the high end, in which substantial evidence is found in favor of the model with applicable priors of $\mathcal{N}(0, 0.5)$ relative to the model with broader priors of $\mathcal{N}(0, 1)$. Pattern uniformity could potentially reflect the influences of both target and contrast uniformity together, but primarily reveals an upper limit to apparent deviations from the population template of phonetic targets.

TABLE 2. Bayes factors of models varying in the specification of relevant prior distributions for testing the strength of the uniformity constraints in the laboratory isolated speech data. The Bayes factor is the ratio between the marginal likelihoods of the data given the specifications for two models, M_1 and M_2. In all cases, M_1 is the model in the top row and M_2, the model in the first column. In any cell, a value greater than 1 indicates evidence in favor of M_1; values less than 1 indicate evidence in favor of M_2. Priors over all fixed effects are presented in TABLE 1 or specified in the sub-caption. Priors over the random by-talker intercept and slopes are implemented as Normal distributions, centered on 0 with a standard deviation of 1 ERB, unless otherwise specified.

a. TARGET UNIFORMITY: POPULATION. Each prior distribution over the fixed effect of [voice] is presented in the header column and row. Each random by-talker intercept and slope has a prior distribution of $\mathcal{N}(0, 0.1)$.

Fixed effect of [voice]	$\mathcal{N}(0, 0.01)$	$\mathcal{N}(0, 0.1)$	$\mathcal{N}(0, 0.5)$	$\mathcal{N}(0, 1)$	$Unif(-10,10)$
$\mathcal{N}(0, 0.01)$		2.57	0.85	0.44	0.06
$\mathcal{N}(0, 0.1)$	0.39		0.33	0.17	0.02
$\mathcal{N}(0, 0.5)$	1.17	3.01		0.52	0.06
$\mathcal{N}(0, 1)$	2.28	5.85	1.94		0.13
$Unif(-10,10)$	18.06	46.43	15.41	7.94	

b. TARGET UNIFORMITY: TALKER. Each prior distribution over the random by-talker slope for voice is presented in the header column and row. The prior over the fixed effect of [voice] is specified as $\mathcal{N}(0, 0.1)$. All other random by-talker effects have a prior distribution of $\mathcal{N}(0, 0.1)$.

Random by-talker slope for [voice]	$\mathcal{N}(0, 0.01)$	$\mathcal{N}(0, 0.1)$	$\mathcal{N}(0, 0.5)$	$\mathcal{N}(0, 1)$
$\mathcal{N}(0, 0.01)$		3.14	1.41	0.73
$\mathcal{N}(0, 0.1)$	0.32		0.45	0.23
$\mathcal{N}(0, 0.5)$	0.71	2.22		0.52
$\mathcal{N}(0, 1)$	1.37	4.29	1.93	

c. CONTRAST UNIFORMITY. Each prior distribution over the random by-talker slopes for [anterior] is presented in the header column and rows. The prior over the fixed effect of [voice] is specified as $\mathcal{N}(0, 0.1)$. All other random by-talker effects have a prior distribution of $\mathcal{N}(0, 0.1)$.

Random by-talker slope for [anterior]	$\mathcal{N}(0, 0.01)$	$\mathcal{N}(0, 0.1)$	$\mathcal{N}(0, 0.5)$	$\mathcal{N}(0, 1)$
$\mathcal{N}(0, 0.01)$		>10,000	>10,000	>10,000
$\mathcal{N}(0, 0.1)$	<0.001		4462.96	3580.95
$\mathcal{N}(0, 0.5)$	<0.001	<0.001		0.80
$\mathcal{N}(0, 1)$	<0.001	<0.001	1.25	

d. PATTERN UNIFORMITY. Prior distributions over the random by-talker slopes for [anterior], [voice] and [anterior] × [voice] are presented in the header column and rows. These priors are specified in the same manner for each random effect. The prior over the fixed effect of [voice] is specified as $\mathcal{N}(0, 0.1)$.

Random by-talker slopes for [anterior], [voice] and [anterior] × [voice]	$\mathcal{N}(0, 0.01)$	$\mathcal{N}(0, 0.1)$	$\mathcal{N}(0, 0.5)$	$\mathcal{N}(0, 1)$
$\mathcal{N}(0, 0.01)$		>10,000	>10,000	>10,000
$\mathcal{N}(0, 0.1)$	<0.001		832.62	179.75
$\mathcal{N}(0, 0.5)$	<0.001	0.001		0.22
$\mathcal{N}(0, 1)$	<0.001	0.006	4.63	

3.2.2.4. Model interpretation

In the final analysis, we report the overall estimates from a linear mixed-effects model with the most credible prior specifications according to the model comparisons. This model has a prior distribution of $\mathcal{N}(0, 0.1)$ for the main effect of [voice] and $\mathcal{N}(0, 0.5)$ for each random by-talker slope; all other prior distributions are specified in TABLE 1. Given the Bayesian framework, we report the beta estimate and 95% credible interval for each effect. Credibility in the direction of an effect is determined based on whether the 95% credible interval excludes zero. For

interpretability, we summarize the model and also report the predicted mean mid-frequency peak values in ERB and hertz in TABLE 3.

Place of articulation, [anterior], had a large and positive effect on mid-frequency peak (β_{place} = 2.32, 95% CrI: [2.08, 2.56]). The effect of [voice] was small and positive, but not reliable in its direction (β_{voice} = 0.08, 95% CrI: [0.00, 0.16]). (Note that the model with a uniform prior over [voice] gave rise to a highly comparable estimate that was only marginally reliable in its positive direction: β_{voice} = 0.10, 95% CrI: [0.01, 0.19].) A reliable interaction was found between [anterior] and [voice], such that the difference in mid-frequency peak between [s] and [ʃ] was somewhat smaller than that between [z] and [ʒ] ($\beta_{place\ x\ voice}$ = -0.10, 95% CrI: [-0.18, -0.01]). The effects of vowel height, vowel backness, and their interaction were also reliable: high vowels corresponded to lower mid-frequency peaks (β_{height} = -0.17, 95% CrI: [-0.28, -0.07]), though this was tempered by a positive interaction between height and backness, which likely reflected a noticeable difference in mid-frequency peak between the high front vowels [i] and [ɪ] and the high, rounded back vowel [u] ($\beta_{height\ x\ backness}$ = 0.19, 95% CrI: [0.09, 0.30]). Following front vowels corresponded to higher sibilant mid-frequency peaks than following back vowels ($\beta_{backness}$ = 0.62, 95% CrI: [0.53, 0.67]). The observed difference between front and back vowels could potentially be explained by the slightly confounded effect of vowel rounding. The mean sibilant mid-frequency peak was numerically lowest preceding the rounded back vowel [u], [o], and [ɔ], but the mean mid-frequency peak before any non-front vowel was indeed lower than the mean mid-frequency peak before any front vowel. Finally, female speakers had a reliably higher mid-frequency peak than male speakers (β_{gender} = 0.38, 95% CrI: [0.14, 0.62]).

TABLE 3. Model estimates and 95% credible intervals for each fixed effect in the linear regression model of sibilant mid-frequency peak in the isolated laboratory speech. The predicted mean mid-frequency peaks for each level of a predictor are also provided in hertz and in ERB.

Predictor	**Model Estimate [95% CrI]** ERB	**Levels**	**Predicted Mean Mid-Frequency Peak** Hz		ERB	
[anterior]	2.32 [2.08, 2.56]	[+anterior] [-anterior]	4800 4675		30.4 25.3	
[voice]	0.08 [0.00, 0.16]	[-voice] [+voice]	4745 4744		28.1 28.1	
[anterior] x [voice]	-0.10 [-0.18, -0.01]	[s] [ʒ] [z] [ʃ]	4800 4801	4674 4678	30.4 30.4	25.2 25.4
following vowel height	-0.17 [-0.28, -0.07]	[+high] [-high]	4747 4742		28.2 28.0	
following vowel backness	0.62 [0.53, 0.67]	[+front] [-front]	4763 4726		28.8 27.4	
vowel height x backness	0.19 [0.09, 0.30]	[+fr, +hi], [-fr, –hi] [+fr, –hi], [-fr, +hi]	4766 4761	4729 4715	29.0 28.8	27.5 26.9
gender	0.38 [0.14, 0.62]	female male	4752 4724		28.4 27.3	

3.3. Discussion

Considerable variation was observed across talker-specific means and standard deviations of mid-frequency peak for each sibilant segment. Variation in the talker-specific means was also moderately to strongly structured between sibilant segments. Strong correlations of talker mean mid-frequency peak were observed between homorganic sibilants, and the paired means were very similar to one another. In contrast, correlations between sibilants contrasting in place of articulation were fairly weak, indicating that the difference between phonetic targets of contrasting features was not consistent across talkers. These correlational findings lend support to a very strong constraint of target uniformity, and a weaker constraint of contrast uniformity.

The strengths of target, contrast, and pattern uniformity were further assessed in the Bayesian analysis. Target uniformity should constrain the mapping from the place of articulation

feature of a sibilant to the corresponding phonetic target, approximated here using the mid-frequency peak; as such, the influence of [voice] on mid-frequency peak should be very minimal, both in the population and within individual talkers. By modulating the breadth of the prior distribution around a null influence of [voice], we ascertained that among the priors tested, the data were most consistent with a prior distribution of $\mathcal{N}(0, 0.1)$ for [voice]. Importantly, the data were much more consistent with this model than with ones containing broader prior distributions over [voice]. The actual estimated effect of [voice] on mid-frequency peak was approximately 0.08 ERB, and in the model predictions of [-voice] and [+voice] mean mid-frequency peaks, the difference was only 1 Hz on average, and a maximum difference of 4 Hz between homorganic sibilants. This is very small, suggesting a very strong constraint of target uniformity in the population. Moreover, talkers did not deviate considerably from this population difference, as indicated by the model comparisons among priors over the random by-talker slope for [voice]. Among priors tested, the data were most consistent with the model containing a prior of $\mathcal{N}(0, 0.1)$ for any additional talker-specific deviations from the main effect of [voice].

Consistent with the correlational analysis, the strength of contrast uniformity was considerably weaker than target uniformity. The data were just narrowly more consistent with a model containing a prior of $\mathcal{N}(0, 0.5)$ over a model with the broadest tested prior of $\mathcal{N}(0, 1)$ for the relevant by-talker slope for [anterior].

Finally, for pattern uniformity, moderate evidence was observed for an upper limit on the overall deviations in the template of mid-frequency peak targets among all four sibilant fricatives: the data were substantially more consistent with a prior of $\mathcal{N}(0, 0.5)$ over random by-talker slopes of [anterior], [voice], and their interaction, relative to broader corresponding prior distributions of $\mathcal{N}(0, 1)$ over those particular random slopes. Importantly, this reveals an upper limit on phonetic variation among sibilants across talkers.

The present study examined sibilant fricatives in isolated productions in highly controlled linguistic and physical environments. We identified particularly strong influences of target uniformity, a very minimal influence of contrast uniformity, and an upper limit on variation in the overall pattern of mid-frequency peak targets. Patterns of variation could easily differ in a more naturalistic environment: to assess this further, we turn now to an analysis of uniformity among American English sibilant fricatives in spontaneous speech productions.

4. Uniformity in spontaneous speech

In addition to the isolated speech style, we also examined the predictions of the uniformity constraints on the phonetic realization of sibilant place of articulation in spontaneous American English speech. Spontaneous speech, as the representative speech style for naturalistic variability, presents a critical test case for assessing the consistency of the patterns of variation and covariation found above. We employed the Buckeye Corpus of Spontaneous Speech, which contains oral interviews from 40 native speakers of American English (Pitt et al. 2005). As the speech is naturally occurring, the relative number of tokens varies across sibilant categories, contexts, and talkers; however, many of these differences could be brought under statistical control in the mixed-effects regression analysis.

4.1. Methods

4.1.1. Corpus description

The Buckeye Corpus contains speech produced by 40 native speakers of American English from the Columbus, Ohio area (Pitt et al. 2005). The talker demographics were counterbalanced for gender and age, such that there were 20 female, 20 male, 20 'young' (under age 30), and 20 'old' (over age 40) talkers; all speakers were white and middle to upper class. Each talker was interviewed in a quiet room for 30 to 60 minutes on current local issues, and was naïve to the true purpose of the recording until after the interview had concluded. The analyzed recordings were sampled at 16 kHz. Word-level and phone-level transcriptions and alignments were provided with the corpus.

4.1.2. Data preparation and acoustic analysis

Word-initial and word-medial prevocalic sibilants in the Buckeye corpus were analyzed with the same acoustic measurements and statistical methods as in the previous experiments. Disfluencies and non-word instances were removed from the analysis. In total, 24,418 sibilants were in the analysis. As expected, [s] was well-represented, whereas [ʒ] was quite rare (see Appendix, TABLE 10 for speaker-specific and total counts).

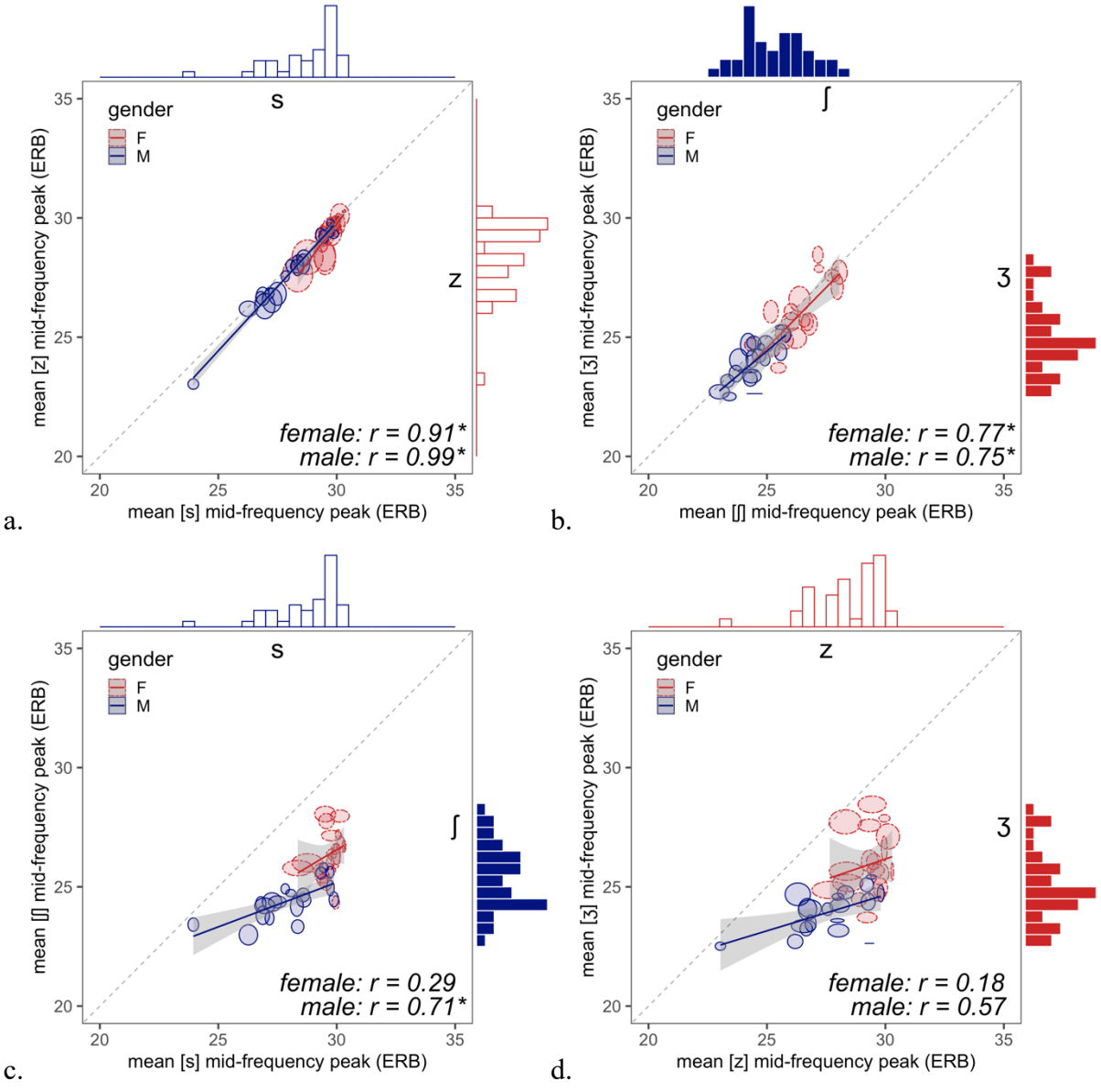
4.2. Results

In the spontaneous speech data, the population means between homorganic sibilants were comparable, and the population standard deviations for [s] and [z] were larger than those for [ʃ] and [ʒ] (Appendix, TABLE 11). Talker means and standard deviations also ranged considerably. Most correlations between talker mean and standard deviations for each sibilant category did not reach significance (Appendix, TABLE 12).

4.2.1. Correlation analysis

As shown in FIGURE 4, strong by-gender correlations of talker mean mid-frequency peak were observed for the [+anterior] sibilants ([s] – [z] female: $r = 0.91$, male: $r = 0.99$, each $p < 0.001$), as well as the [-anterior] sibilants ([ʃ] – [ʒ] female: $r = 0.77$, male: $r = 0.75$, $p < 0.001$; see also Appendix, TABLE 13). Talker mean mid-frequency peak was weakly correlated between the voiceless fricatives [s] and [ʃ] across female speakers (female: $r = 0.29$, $p > 0.01$) and strongly correlated across male speakers (male: $r = 0.71$, $p < 0.001$). Talker mean mid-frequency peak was weakly to moderately correlated between the voiced fricatives [z] and [ʒ] (female: $r = 0.18$, male: $r = 0.57$, each $p > 0.01$).

FIGURE 4. Variation and covariation of sibilant mid-frequency peak (ERB) across talkers in the American English spontaneous speech from the Buckeye Corpus. Each ellipsoid is centered on a pair of talker-specific means and is color-coded by talker gender; the size of the ellipsoid reflects 1/5 of the standard deviation of the respective sibilants. Marginal histograms indicate the variation in talker means for each sibilant category. The asterisk indicates $p < 0.01$. Gray shading reflects the local confidence interval around the best-fit linear regression of talker means for each gender.



4.2.2. Bayesian analysis

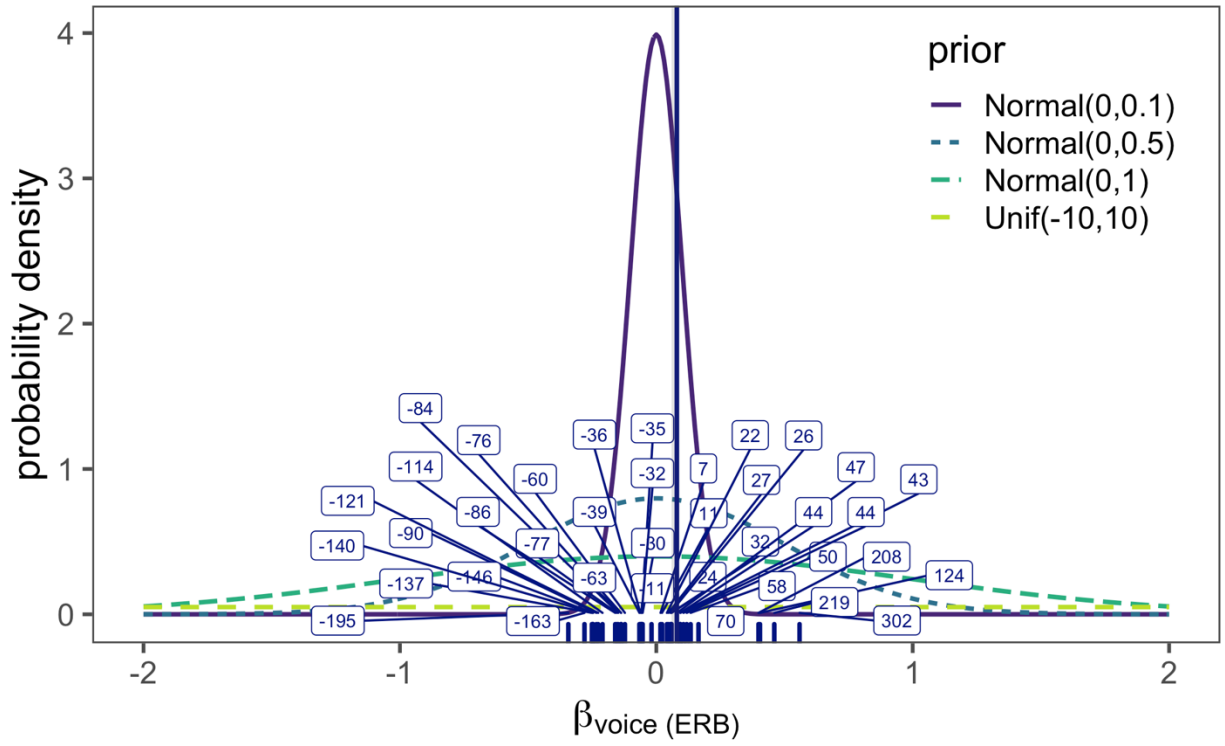
In the Bayesian analysis, we examined the influence of each uniformity constraint on the phonetic realization of sibilant place of articulation, in the same manner as in Section 3. Uniformity constraints were modeled as prior distributions over relevant factors in a linear mixed-effects regression predicting sibilant mid-frequency peak in ERB; differing strengths of each prior were then compared using Bayes factors, which represent ratios between M_1 and M_2. Higher Bayes factors indicate greater evidence towards M_1; see Section 3.2.2 for a description of the Jeffreys' scale interpretation.

4.2.2.1. Target uniformity

In the first set of model comparisons, we manipulated the prior distribution over the effect of [voice] on sibilant mid-frequency peak. A depiction of these prior distributions on the [voice] contrast is shown in FIGURE 5 along with the mean mid-frequency peak deviation of the [-voice] sibilants from the talker-specific mean. As shown in TABLE 4a, extreme evidence exists against the tightest tested prior over [voice], but moderate to very strong evidence is found in favor of the next tightest prior over [voice], $\mathcal{N}(0, 0.1)$, relative to broader priors.

In the second set of model comparisons, we manipulated the prior distribution over the random by-talker slope for [voice]. As shown in TABLE 4b, substantial evidence exists in favor of a model with a prior of $\mathcal{N}(0, 0.1)$ relative to the narrow prior of $\mathcal{N}(0, 0.01)$, and critically to wider prior distributions. This suggests that variation in the strength of target uniformity is minimal, but also present across talkers.

FIGURE 5. Priors over the population effect of [voice]. Given the coding scheme, the prior reflects the distance of the [-voice] mid-frequency peak from the mean. (The effect of [voice] was weighted effect coded: [-voice] = +1, [+voice] = -1.04.) The tightest prior of $\mathcal{N}(0, 0.01)$ is not pictured here due to its concentrated probability density. The rug plot corresponds to half the difference between by-talker [-voice] and [+voice] mid-frequency peak means in ERB in the American English spontaneous speech from the Buckeye corpus. These are labeled with their corresponding contrast in hertz. (Actual by-talker differences range from -391 Hz to 604 Hz across talkers.) The vertical line reflects the estimated mean effect of [voice], 0.08, using the model reported in Section 4.2.2.4. The gray shading represents the 95% credible interval around that estimate ([0.06, 0.09]).



4.2.2.2. Contrast uniformity

In the third set of model comparisons, we investigated the strength of contrast uniformity on mid-frequency peak by modulating the prior distribution over the random by-talker slope for [anterior]. As shown in TABLE 4c, the top model has a prior of $\mathcal{N}(0, 0.5)$, but with only anecdotal evidence in its favor relative to a model with a wider standard deviation of 1. This

suggests that talkers can deviate a fair amount in the mid-frequency peak contrast between [+anterior] and [-anterior] sibilants.

4.2.2.3. Pattern uniformity

In the fourth set of model comparisons, we examine the strength of pattern uniformity in the instantiation of mid-frequency peak across talkers by modulating the priors over all random by-talker slopes. As shown in TABLE 4d, extreme evidence exists against the model with the tightest constraints on cross-talker variation relative to the models with broader standard deviations; however, strong to extreme evidence exists in favor of the model with priors of $\mathcal{N}(0, 0.1)$ relative to models with larger standard deviations. This suggests a reasonably high degree of consistency in the overall template of mid-frequency peak across talkers.

TABLE 4. Bayes factors of models varying in the specification of relevant prior distributions for testing the strength of the uniformity constraints in the American English spontaneous speech from the Buckeye Corpus. The Bayes factor is the ratio between the marginal likelihoods of the data given the specifications for two models, M_1 and M_2. In all cases, M_1 is the model in the top row and M_2, the model in the first column. Values greater than 1 indicate evidence in favor of M_1; values less than 1 indicate evidence in favor of M_2. Priors over all fixed effects are presented in TABLE 1 or specified in the sub-caption. Priors over the random by-talker intercept and slopes are implemented as Normal distributions, centered on 0 with a standard deviation of 1 ERB, unless otherwise specified.

a. TARGET UNIFORMITY: POPULATION. Each prior distribution over the fixed effect of [voice] is presented in the header column and row. Each random by-talker intercept and slope has a prior distribution of $\mathcal{N}(0, 1)$.

Fixed effect of [voice]	$\mathcal{N}(0, 0.01)$	$\mathcal{N}(0, 0.1)$	$\mathcal{N}(0, 0.5)$	$\mathcal{N}(0, 1)$	Unif$(-10, 10)$
$\mathcal{N}(0, 0.01)$		>10,000	>10,000	>10,000	>10,000
$\mathcal{N}(0, 0.1)$	<0.001		0.27	0.14	0.02
$\mathcal{N}(0, 0.5)$	<0.001	3.73		0.53	0.06
$\mathcal{N}(0, 1)$	<0.001	7.09	1.90		0.12
Unif$(-10, 10)$	<0.001	58.6	15.72	8.27	

b. TARGET UNIFORMITY: TALKER. Each prior distribution over the random by-talker slope for [voice] is presented in the header column and row. The prior over the fixed effect of [voice] is specified as $\mathcal{N}(0, 1)$. All other random by-talker effects have a prior distribution of $\mathcal{N}(0, 1)$.

Random by-talker slope for [voice]	$\mathcal{N}(0, 0.01)$	$\mathcal{N}(0, 0.1)$	$\mathcal{N}(0, 0.5)$	$\mathcal{N}(0, 1)$
$\mathcal{N}(0, 0.01)$		9.27	2.02	1.00
$\mathcal{N}(0, 0.1)$	0.11		0.22	0.11
$\mathcal{N}(0, 0.5)$	0.49	4.58		0.49
$\mathcal{N}(0, 1)$	1.00	9.30	2.03	

c. CONTRAST UNIFORMITY. Each prior distribution over the random by-talker slopes for [anterior] is presented in the header column and rows. The prior over the fixed effect of [voice] is specified as $\mathcal{N}(0, 1)$. All other random by-talker effects have a prior distribution of $\mathcal{N}(0, 1)$.

Random by-talker slope for [anterior]	$\mathcal{N}(0, 0.01)$	$\mathcal{N}(0, 0.1)$	$\mathcal{N}(0, 0.5)$	$\mathcal{N}(0, 1)$
$\mathcal{N}(0, 0.01)$		>10,000	>10,000	>10,000
$\mathcal{N}(0, 0.1)$	<0.001		1.37	0.72
$\mathcal{N}(0, 0.5)$	<0.001	0.73		0.52
$\mathcal{N}(0, 1)$	<0.001	1.39	1.91	

d. PATTERN UNIFORMITY. Prior distributions over the random by-talker slopes for [anterior], [voice], and [anterior] × [voice] are presented in the header column and rows. These priors are specified in the same manner for each random effect. The prior over the fixed effect of [voice] is specified as $\mathcal{N}(0, 1)$.

Random by-talker slopes for [anterior], [voice], and [anterior] × [voice]	$\mathcal{N}(0, 0.01)$	$\mathcal{N}(0, 0.1)$	$\mathcal{N}(0, 0.5)$	$\mathcal{N}(0, 1)$
$\mathcal{N}(0, 0.01)$		>10,000	>10,000	>10,000
$\mathcal{N}(0, 0.1)$	<0.001		0.06	0.01
$\mathcal{N}(0, 0.5)$	<0.001	16.98		0.13
$\mathcal{N}(0, 1)$	<0.001	126.95	7.48	

4.2.2.4. Model interpretation

Finally, we report the estimates of the marginal posterior distributions for each effect in a linear mixed-effects model that has credible prior parameters according to the model comparisons. Note that for the spontaneous speech data, strong evidence was found in favor of a model with priors of $\mathcal{N}(0, 0.1)$ on random by-talker slopes relative to one with priors having broader

standard deviations. However, for consistency with the isolated speech analysis, the reported model has a prior distribution of $\mathcal{N}(0, 0.1)$ for the main effect of [voice] and $\mathcal{N}(0, 0.5)$ for each random by-talker slope; all other prior distributions are specified in TABLE 1. For interpretability, we summarize the model and also report the predicted mean mid-frequency peak values in ERB and hertz in TABLE 5.

Place of articulation had a large and positive effect on mid-frequency peak, though somewhat smaller in magnitude than in the isolated speech model (β_{place} = 0.68, 95% CrI: [0.61, 0.74]). As before, the effect of [voice] was small, positive, and very similar in magnitude; however, in this model, the direction was reliable with a tight posterior distribution (β_{voice} = 0.08, 95% CrI: [0.06, 0.09]). (Note that the model with a uniform prior from -10 to +10 ERB on [voice] gave rise to the exact same estimate and 95% credible interval: β_{voice} = 0.08, 95% CrI: [0.06, 0.09].) Unlike in the isolated speech model, the interaction between [anterior] and [voice] was nonexistent ($\beta_{place\ x\ voice}$ = 0.00, 95% CrI: [-0.01, 0.01]). While the effect of vowel height was not reliable (β_{height} = -0.01, 95% CrI: [-0.05, 0.03]), the effects of vowel backness and the interaction between height and backness were reliable: following front vowels corresponded to slightly higher sibilant mid-frequency peaks than following back vowels ($\beta_{backness}$ = 0.16, 95% CrI: [0.14, 0.19]). The reliable interaction once again likely reflected a noticeable difference in mid-frequency peak between the high front vowels [i] and [ɪ] and the high, rounded back vowel [u] ($\beta_{height\ x\ backness}$ = 0.07, 95% CrI: [0.04, 0.11]). In the present corpus, the observed difference between front and non-front vowels did indeed reflect the difference in frontness as opposed to rounding. Though all front vowels were unrounded, there was a reasonable balance between unrounded and rounded non-front vowels. The mean mid-frequency peaks per sibilant between the non-front rounded and unrounded variants were not reliably different from one another. Finally, female speakers had a reliably higher mid-frequency peak than male speakers (β_{gender} = 1.05, 95% CrI: [0.75, 1.35]).

TABLE 5. Model estimates and 95% credible intervals for each fixed effect in the linear regression model of sibilant mid-frequency peak in the American English spontaneous speech from the Buckeye Corpus. The predicted mean mid-frequency peaks for each level of a predictor are also provided in hertz and in ERB.

Predictor	**Model Estimate [95% CrI]** ERB	**Levels**	**Predicted Mean Mid-Frequency Peak**			
			Hz		ERB	
[anterior]	0.68 [0.61, 0.74]	[+anterior] [-anterior]	4678 4602		28.7 25.3	
[voice]	0.08 [0.06, 0.09]	[-voice] [+voice]	4663 4663		28.1 28.1	
[anterior] x [voice]	0.00 [-0.01, 0.01]	[s] [ʒ] [z] [ʃ]	4680 4671	4594 4603	28.8 28.5	24.9 25.4
following vowel height	-0.01 [-0.05, 0.03]	[+high] [-high]	4655 4667		27.7 28.3	
following vowel backness	0.16 [0.14, 0.19]	[+front] [-front]	4665 4661		28.1 28.0	
vowel height x backness	0.07 [0.04, 0.11]	[+fr, +hi], [-fr, –hi] [+fr, –hi], [-fr, +hi]	4660 4672	4665 4622	27.9 28.4	28.2 26.2
gender	1.05 [0.75, 1.35]	female male	4683 4643		28.9 27.2	

4.3. Discussion

In spontaneous speech, talkers varied substantially in their realization of sibilant mid-frequency peak. Consistent with previous findings, the overall standard deviation across talkers was higher here than in the isolated speech style, as was the range of talker-specific standard deviations. Patterns of structured variation among sibilants nevertheless emerged in a way that closely mirrored the patterns in isolated speech. Correlations of talker mean mid-frequency peak were very strong between sibilants with a shared place of articulation, whereas correlations between sibilants contrasting in place of articulation were weak to moderate, though these reached significance across male talkers. The former findings lend strong support towards target

uniformity, and the latter findings suggest a potential role of contrast uniformity in shaping phonetic structure.

By modulating the relevant prior distributions, we could also identify upper and lower bounds on the prior influences of target, contrast, and pattern uniformity on sibilant mid-frequency peak. With respect to target uniformity, the resulting influence of [voice] was found to be 0.08 ERB, which is the same estimate of [voice] found in the isolated speech data with different talkers. The model predicted no difference between the overall means of [-voice] and [+voice] sibilant mid-frequency peak. Moreover, the data were most consistent with models that placed a reasonably tight constraint on the population influence of [voice] both within and across speakers. Substantial evidence was found in favor of a model with a prior of $\mathcal{N}(0, 0.1)$ on the main effect of [voice] and a prior of $\mathcal{N}(0, 0.1)$ on the random by-talker slope for [voice]. These findings support a strong constraint that minimizes the influence of [voice] on sibilant mid-frequency peak.

With respect to contrast uniformity, anecdotal evidence was found in favor of the model with a prior over the random by-talker slope for [anterior] of $\mathcal{N}(0, 0.5)$. This model did not differ considerably from one with a broader prior of $\mathcal{N}(0, 1)$ of the same effect. This pattern of findings is highly comparable to that found in the isolated speech data. Considerable variation is thus found across talkers in the overall difference between [+anterior] and [-anterior] sibilants, suggesting a very weak constraint of contrast uniformity. With respect to pattern uniformity, an upper limit on overall variation in the population was nevertheless identified. Strong to decisive evidence exists in favor of a model with priors of $\mathcal{N}(0, 0.1)$ over the random by-talker slopes of [anterior], [voice], and their interaction relative to models with broader standard deviations on those same effects. This particular model exceeds the influence of target uniformity alone: not only does the strongest model have a random by-talker slope for [voice] prior of $\mathcal{N}(0, 0.1)$, but so do the random by-talker slopes of [anterior] and the interaction between [anterior] and [voice].

5. General Discussion

In the present paper, we investigated the extent of variability and systematicity in the phonetic targets for place of articulation in sibilant fricatives, both within and across speakers of American English and in different speech styles. We assessed the viability and strength of three

constraints on the phonetics–phonology interface that could restrict variation in this phonetic space: target, contrast, and pattern uniformity.

A Bayesian linear mixed-effects model was used to model the mapping from distinctive features to phonetic targets and their physical instantiations with a rich by-talker random effect structure. We assumed this discrete grouping variable was the feature [anterior], though we could have alternatively used the label [alveolar] (or even [X]). Additionally, we assumed a likely phonetic target present in sibilant fricatives would be a phonetic place of articulation specification with integrated articulatory and perceptual targets. These considerations motivated our choice of the spectral mid-frequency peak in ERB as the operationalization of the phonetic target: the mid-frequency peak broadly reflects the articulatory place of articulation via the front cavity resonance, and ERB provides a perceptual scaling of the frequency spectrum.

The three uniformity constraints were then modeled as prior distributions that constrain the mapping from the phonological segment to the corresponding phonetic targets within and across talkers. An additional correlational analysis assessed the strength of the talker mean mid-frequency peak relationships among sibilant fricatives. Pattern uniformity places constraints on this mapping without reference to the internal structure of a segment, whereas target and contrast uniformity require a discrete internal representation.

In the following sections, we first present a summary of the present findings with respect to target, contrast, and pattern uniformity. We then examine how uniformity may account for previous observations of phonetic structure, and discuss the implications of phonetic structure and uniformity for language variation and change, acquisition, and perceptual adaptation.

5.1. Summary

Variation in sibilant mid-frequency peak was considerable across talkers and highly structured. In particular, the correlational and Bayesian analyses supported a very strong constraint of target uniformity on the phonetic realization of sibilant fricatives. Talker mean mid-frequency peaks were strongly correlated between homorganic sibilant fricatives, and also very similar to one another. The influence of the [voice] feature on mid-frequency peak was also minimal in both the isolated and spontaneous speech data: based on the model predictions, almost no difference was found between the mid-frequency peak means of voiced and voiceless sibilants in both corpora. Importantly, it seems highly probable that a speaker would be physically able to produce a larger

contrast in mid-frequency peak, should they be so inclined. Strong evidence was found in favor of tight prior distributions over the fixed effect and random by-talker slope for [voice] with $\mathcal{N}(0, 0.01)$ to $\mathcal{N}(0, 0.1)$ for the main effect, and $\mathcal{N}(0, 0.1)$ for the random by-talker slope.

In comparison to target uniformity, evidence for contrast uniformity was relatively weak in both studies. Despite some significant correlations of talker mean mid-frequency peak between sibilants contrasting in anteriority, the Bayesian analysis revealed only anecdotal evidence in favor of an upper limit on talker-specific deviations from the estimated population contrast in both speech styles. In other words, talkers modulated the size of the contrast between [+anterior] and [-anterior] contrasts considerably.

Despite the weak support for contrast uniformity in the present analyses, an upper limit on overall variation in the template was apparent in the present studies. In the isolated speech data, the variation was somewhat broader; moderate evidence was found in favor of prior distributions specified as $\mathcal{N}(0, 0.5)$ relative to $\mathcal{N}(0, 1)$ on the random by-talker slopes. In the spontaneous speech data, moderate to very strong evidence was found in favor of prior distributions specified as $\mathcal{N}(0, 0.1)$ relative to ones with broader standard deviations on the random by-talker slopes. Taken together, talker-specific deviations from the population template may be limited to standard deviations of 0.1 to 0.5 ERB.

In both speech styles, the [anterior] specification had the largest effect on variation in mid-frequency peak, in line with expectations; the contrast was much larger in the isolated speech style than in the spontaneous speech style. (Using the model predicted means, the [anterior] difference is 125 Hz in the isolated speech and 76 Hz in the spontaneous speech). As reported above, the effect of [voice] on mid-frequency peak was very small and not reliable in its direction for the laboratory speech data. A reliable interaction was observed between [anterior] and [voice] in the isolated speech data, but not the spontaneous speech data. In both speech styles, the effect of vowel height was not reliable, whereas the effect of vowel backness and the interaction between vowel height and backness were reliable. Front vowels corresponded to slightly higher mid-frequency peaks than back vowels, and based on observations from the spontaneous speech data, this effect did at least in part arise from a difference in backness, as opposed to the partially confounded contrast in rounding. The interaction between height and backness indicated an even larger contrast in mid-frequency peak between front vowels [i] and [ɪ] and the back vowel [u] than would be expected based on the independent specifications of

height and backness alone. Finally, female speakers had a reliably higher sibilant mid-frequency peak than male speakers, and despite the apparent difference in the model estimates between isolated and spontaneous speech, the estimated average difference was only slightly higher for spontaneous than isolated speech. (Using the model predicted means, sibilant mid-frequency peaks were on average 28 Hz higher for female than male speakers in the isolated speech data, and 40 Hz higher in the spontaneous speech data. Note that these estimates already take into account individual variation in the population via the random by-talker effects.)

5.2. Target uniformity

The current findings lend support to a strong constraint of target uniformity on the phonetic realization of sibilant fricative place of articulation, as measured by the mid-frequency peak. However, several findings from the literature, concerning stop VOT, intrinsic vowel f0, and intrinsic vowel duration could be interpreted as ostensibly contradicting target uniformity. We summarize these cases below. Under proper characterization of the phonetic targets, these cases can be seen to support rather than undermine the proposed constraint.

5.2.1. Stop consonant voice onset time

For stop consonants with a shared laryngeal status, VOT is inversely related to place of articulation: stops with voicing lead generally decrease in duration with more posterior places of articulation (/b/ > /d/ > /g/), whereas stops with voicing lag generally increase in duration with more posterior places of articulation (/p/ < /t/ < /k/; Maddieson 1997; Cho and Ladefoged 1999). The acoustic correlate to the stop laryngeal feature thus differs across stops within a laryngeal series. The slight difference in the acoustic measurement, though, can be straightforwardly accounted for by a uniform realization of the glottal spreading gesture and its timing relative to the oral constriction (Maddieson 1997). As further support for this, strong covariation of talker mean VOT has been found among aspirated stop consonants in American English (Chodroff and Wilson 2017) and German (Hullebus et al. 2018). Chodroff and Wilson (2017) reported Pearson correlations of talker means that were at or above 0.95 in a 24-talker corpus of isolated speech and above 0.75 in a 180-talker corpus of connected speech. Across the aspirated stop consonants, the talker mean VOTs were also highly comparable, but generally increased in duration with more posterior places of articulation. These findings strongly implicate a near-uniform

realization of the shared phonological feature underlying VOT (e.g. [+spread glottis]) within a talker.

Previous studies on English aspirated stops have, however, reported variation in the relative ordering of [tʰ] and [kʰ] (e.g. Docherty 1992; Yao 2009; Chodroff and Wilson 2017). Articulatory evidence from English production indicates a longer glottal opening gesture for [tʰ] than for [kʰ], suggesting that the phonetic target for the [+spread glottis] feature is *not* uniform for each segment (Cooper 1991; Hoole and Pouplier 2015).[9] Instead, the presence of [CORONAL] appears to interact with the duration of the glottal spreading gesture. The [CORONAL] feature has a relatively unmarked status, and coronals may enjoy somewhat greater freedom of phonetic realization than other segments. Nevertheless, the observed variation is minimal, especially considering the otherwise consistent cross-linguistic patterns. While there may be some context sensitivity between [spread glottis] and [CORONAL], the overall patterns suggest a strong influence of target uniformity on the phonetic realization of [spread glottis].

5.2.2. Intrinsic vowel f0

On the surface, intrinsic vowel f0 may also appear to reflect a weak influence of target uniformity. Intrinsic f0 refers to the cross-linguistic observation that the fundamental frequency (f0) of high vowels such as [i] and [u] is higher than that of low vowels such as [a] (e.g. Mohr 1971; Whalen and Levitt 1995). One explanation for intrinsic f0 is that high and low vowels have different phonetic targets for the rate of vocal fold vibration. The existence of tone languages, however, shows that maintaining the opposite relationship between high and low vowels is physically possible: high tones can exist on low vowels, just as low tones can exist on high vowels. An alternative explanation, suggested by the term ‘intrinsic f0’, is that pitch differences arise from an interaction between the articulations of tongue height and voicing. Though the precise articulatory specifications are debated, the raised tongue body of high vowels could consequently raise the hyoid bone, causing increased tension on the laryngeal system, and

[9] Deviations from the presumed universal ranking have also been observed in Dahalo and Navajo (Cho and Ladefoged 1999). In Dahalo, the average VOT for unaspirated [t] was greater than the VOT for [k] ([t]: 42 ms, [k]: 27 ms), and in Navajo, the VOT for unaspirated [t] was lower than both unaspirated [p] and [k] ([p]: 12 ms, [t]: 6 ms, [k]: 45 ms).

thus, a higher f0 (e.g. Lehiste 1970; Ohala 1972). The difference in f0 between high and low vowels may therefore by an automatic consequence of a uniform voice target for f0, but where the tongue height subsequently increases laryngeal tension and raises f0.

Though the difference in f0 between high and low vowels is small (approximately 4 to 25 Hz; Ohala and Eukel 1987), the intrinsic f0 effect is consistent across speakers and languages (Whalen and Levitt 1995). A re-analysis of the Whalen and Levitt (1995) f0 data revealed that language-specific f0 means for [i] and [u] were not only consistently higher than those for [a], but also strongly correlated with each $r > 0.98$. Some languages may nevertheless deviate from exactly uniform laryngeal targets, but the observed differences between languages are consistently small. The stable direction of the relationship and the strong correlation indicates a strong pressure to maintain a high degree of similarity between the laryngeal settings for the high and low vowels. Altogether, this suggests a strong influence of target uniformity on the phonetic realization of the [voice] feature in vowels.

5.2.3. Intrinsic vowel duration

Intrinsic vowel duration may be another candidate for ostensibly violating target uniformity, at least assuming an acoustic target. Across languages, low vowels such as [a] have longer durations than high vowels such as [i] or [u] (e.g. Lindblom 1967). Low vowels require greater articulatory movement in jaw opening, and the durational difference could still arise from a uniform phonetic target for duration (Lindblom 1967). In an articulatory study of vowel production, Westbury and Keating (1980) found that force input to the jaw had not only greater amplitude but also longer duration for [a] than [i]. The fact that talkers accentuate force input beyond that required for an acoustic difference in duration suggests that these two vowels do not share a single target on the dimensions relevant for duration, but rather that the target is affected by the segment-internal [high] and [low] specifications. This could be indicative of a weaker influence of target uniformity. Nevertheless, an analysis of intrinsic vowel duration conducted on a corpus of 180 speakers of American English revealed strong correlations of the pattern of vowel durations across talkers (median $r = 0.90$; Wilson and Chodroff 2017). Moreover, the differences in duration between vowels of different heights were considerably smaller than those found between vowels at different speaking rates or even between short and long vowels in languages with length contrasts (Johnson and Martin 2001). The strong correlations and small

between-vowel differences minimally suggest an influential role of pattern uniformity in the phonetic targets for vowel duration.

5.2.4. Additional cases

A few additional cases consistent with a strong influence of target uniformity are discussed here. Stability in the realization of vowel height, as measured by F1, has been found for vowels in American and British English, Dutch, European and Canadian French, Japanese, European and Peruvian Spanish, and European and Brazilian Portuguese (Watt 2000; Ménard et al. 2008; Oushiro 2019; Schwartz and Ménard 2019). Ménard et al. (2008) further demonstrate via articulatory simulation that this vowel F1 stability is likely generated by a highly consistent tongue height.

Similarly, Faytak (2018) observed a high degree of within-talker systematicity in the precise tongue posture used for fricative vowels with a postalveolar constriction, as well as alveolopalatal fricative consonants in Suzhou Chinese. The author argues that speakers reuse this articulation uniformly with only some idiosyncratic deviations. Overall, these findings are consistent with target uniformity in the phonetic realization of a shared phonological primitive of these segments.

In an articulatory and acoustic analysis of oral and nasal vowels in American English, Carignan et al. (2011) identified a near-uniform F1 across the high vowels [i] and [ĩ], but with different tongue configurations. For the low vowels, [a] and [ã], the tongue height was very consistent, but F1 differed. One explanation is that the F1 similarity between [i] and [ĩ] could prevent neutralization of the perceptual contrast with the neighboring high lax vowel [ɪ], whereas the non-acoustic uniformity of F1 (but articulatory uniformity) for [a] and [ã] does not endanger any perceptual contrast. The complexity of this case presents an interesting opportunity to further explore the nature of phonetic targets (see Section 5.5) and also potential interactions with additional constraints that may structure the phonetic space of a speaker, such as perceptual distinctiveness.

5.3. Contrast uniformity

Evidence for contrast uniformity was quite weak in this study for sibilant place of articulation, as well as in Chodroff and Wilson (2017) for American English stop consonant voicing, as

measured by positive VOT. Correlations of talker-specific mean VOT between homorganic stops ranged from $r = 0.18$ to $r = 0.33$ in the isolated speech study and from $r = 0.15$ to $r = 0.53$ in the connected speech study. Correlations involving the short-lag (or phonologically voiced) stops were thus considerably weaker than those among the long-lag stops. The failure to observe contrast uniformity in these cases could be due to the variable realization of voicing (e.g. variable utilization of negative VOT; Davidson 2016), or simply indicate that contrast uniformity does not constrain phonetic realization. The minimal covariation that was observed among sibilant mid-frequency peak and stop consonant VOT could reduce to trade-offs between phonetic dispersion and articulatory ease. Talkers might achieve sufficient dispersion or distinctiveness among speech sounds that contrast in anteriority (or voicing), after which they are free to vary according to ease of articulation and idiolectal preferences.

Some evidence for contrast uniformity has, however, been observed in Japanese stop VOT. Tanner et al. (2019) found strong systematicity among Japanese stops contrasting in voice, just as contrast uniformity would predict. This is consistent with several scenarios: first, contrast uniformity may be a universal constraint, but with a tendency to have very weak influence across languages. Some languages may prioritize contrast uniformity more than others while structuring phonetic targets. So far, however, it appears that evidence for target uniformity is consistently strong and stronger than for contrast uniformity.

5.4. The nature of phonetic targets

A key component of the phonetics–phonology interface described here is the set of phonetic targets corresponding to individual speech sounds. The nature of these targets has been subject to considerable discussion in the literature. In the present study, we assumed an integrated auditory-articulatory target for sibilant place of articulation. Previous research suggests this resonant frequency should reflect the oral cavity anterior to the tongue constriction (Koenig et al. 2013); we then added an ERB transformation to approximate listener perception.

Perceptuomotor targets seem to be reasonable representations underlying speech production (Guenther 1994, 1995; Schwartz et al. 2007; Ménard et al. 2008; Ghosh et al. 2010). Previous research suggests speakers modulate their phonetic targets based on feedback from auditory and articulatory perturbations, but with individual differences in the dominant feedback mode (Lametti et al. 2012). However, whether the phonetic targets integrate perceptual and

motor targets into one dimension, or whether distinct phonetic targets exist for articulatory and auditory dimensions remains open for further investigation. The nature of the phonetic target representation could even differ across individual speakers. Such tradeoffs between articulatory and acoustic uniformity have also been observed for oral and nasal vowels (Shosted et al. 2012; Carignan 2013) and stop voicing (Keating 2003); in some cases, the phonetic target is more clearly articulatory rather than auditory-acoustic (e.g. Chodroff and Wilson 2017; Faytak 2018). Further research is necessary to determine the nature of phonetic targets that uniformity may restrict, and the additional phonetic constraints (e.g. perceptual distinctiveness) that may account for that selection.

5.5. Implications for language variation and change, acquisition, and perceptual adaptation

Phonetic uniformity has several implications for language variation and change, native and non-native language acquisition, and perceptual adaptation and cross-segment generalization. Target uniformity places strict constraints on the social expressivity of language. As evidenced by the present findings, certain phonetic targets are yoked together across segments by virtue of a shared feature. While the precise phonetic realization of [s] has been shown to convey a range of socioindexical properties, the phonetic specification of place for [s] should immediately constrain the phonetic specification of place for [z] and vice versa. General linguistic constraints on social expressivity in language has previously been discussed as a type of linguistic coherence, in which variability is limited by general structural or grammatical constraints (Guy 2013; Guy and Hinskens 2016). Uniformity may simply be a specific instance of such linguistic coherence.

In the sense that uniformity constrains social linguistic expression, target uniformity may also relate to the notion of parallel shifts in sound change (Fruehwald 2013, 2017). Comparable to the proposal presented here, Fruehwald (2013) posits that changes in the phonetic targets of multiple segments may be governed by a single change in a shared underlying phonological feature that results in a parallel phonetic shift of the natural class. Documented parallel shifts include back vowel fronting (Fridland 2001; Haddican et al. 2013; Labov et al. 2013; Labov 2014) and mid vowel raising (Watt 2000), though the degree to which this holds in sound change more generally may be mixed. Fruehwald (2019) found phonological grounding for changes in the frontness of back vowels in apparent time, and lack of parallelism among other less

featurally-related vowels. In a study of mid-vowel raising across Brazilian migrant speakers in São Paolo, Oushiro (2019) identified some evidence for parallelism in the degree to which speakers modulate both [e] and [o] to match the ambient dialect. These mixed findings provide further examples for some limitations of uniformity and a potential outranking of uniformity by alternative constraints.

As a constraint on the phonetics–phonology interface, target uniformity is also expected to influence feature–target pairings universally. In support of this, covariation of VOT was observed among stop consonants with a shared laryngeal specification in over 100 languages from 36 language families (Chodroff et al. 2019). Moreover, strong covariation of language-specific means was observed not only among aspirated voiceless stop consonants, or long-lag VOT, but also among stop consonants with short-lag and lead VOT. This finding highlights the tight similarity in the language-specific phonetic targets of stop consonants with a shared laryngeal specification. Additional findings from Salesky et al. (2020) identified significant covariation of mean F1 between mid vowels [e] and [o] across 35 typologically diverse languages ($r = 0.62$), and between high vowels [i] and [u] across 40 typologically diverse languages ($r = 0.79$). The correlation of mid-frequency peak (defined as the peak frequency between 3000 and 7000 Hz) between [s] and [z] was also strong and significant across 18 diverse languages ($r = 0.86$). Further research is necessary to determine whether uniformity applies universally in the realization of sibilant place of articulation and other feature–target pairings.

In addition, uniformity has several implications for child and non-native language acquisition. In acquisition generally, target uniformity may allow for a type of bootstrapping between segments in production and perception. If a speaker masters the phonetic targets of a frequently occurring sound (e.g. [s] or [ʃ]), the uniform aspects of production may transfer to a second, less frequent sound (e.g. [z] or [ʒ]). This could extend to other rare sounds with more frequent counterparts as well. Indeed, evidence for target uniformity has been observed in the speech of children as young as four years old for vowel F1 across vowels with a shared height feature (Ménard et al. 2008). Whether such systematic relations are learned via exposure or via an innate uniformity constraint remains open to investigation.

With respect to non-native language acquisition, speakers must learn novel phonetic representations for the target language. This could potentially be done segment-by-segment, or in accordance with target uniformity as a natural class. Chodroff and Baese-Berk (2019) found

evidence that L2 speakers of English minimally maintain the same relationship of VOT between voiceless stop categories, even though the absolute value was not always native-like. Preliminary research also suggests that L2 speakers of English also shift the phonetic targets underlying VOT from their native language to their non-native English speech as a natural class. That is, L2 speakers do indeed change the representation underlying VOT for English, and they make this change not just for one or two segments, but for the whole natural class.

Finally, listeners could exploit structured variation that arises from uniformity to generalize talker-specific phonetic properties from one speech sound to another in rapid adaptation (see Chodroff and Wilson 2020, for related discussion and investigations). Perceptual evidence in support of such generalization across segments has been found for stop consonant VOT (e.g. Kraljic and Samuel 2006; Theodore and Miller 2010) and vowel F1 (e.g. Maye et al. 2008). Listeners could either have knowledge of uniformity or simply exploit the empirical covariation present in the language; regardless, uniformity minimally gives rise to the presence of systematic relationships.

6. Conclusion

Variation in speech within and across individual talkers is substantial, and also highly structured among speech sounds. In the present study, we investigated a set of uniformity constraints that may constrain cross-talker and cross-segment variation in the phonetic realization of speech sounds, with a focus on the realization of sibilant place of articulation, as approximated by the mid-frequency peak (ERB). The primary constraints considered were target uniformity, which requires uniform realization of a shared phonological primitive within a talker; contrast uniformity, which requires a uniform contrast in phonetic realization between segments differing in a phonological primitive across talkers; and pattern uniformity, which requires a uniform template in the phonetic realization of differing segments across talkers.

Using a correlation analysis and a Bayesian hierarchical model of spectral mid-frequency peak, we evaluated the strength of each of these constraints. Strong covariation of talker mean mid-frequency peak was observed among sibilants with a shared place of articulation, while covariation of talker means was quite weak between sibilants contrasting in place of articulation. These findings are consistent with a strong influence of target uniformity and a weak to nonexistent influence of contrast uniformity. The pattern of results in the Bayesian analysis

indicated a strong prior influence of target uniformity, a weak influence of contrast uniformity, and a moderate influence of pattern uniformity, indicating an upper limit on the overall talker-specific deviations from the population template. The present findings minimally suggest a reliable and strong influence of target uniformity, demonstrating that shared phonological feature specifications can imply similar phonetic realizations. The discrete primitives of phonology and the continuous targets of phonetics are found to be more tightly yoked than previously recognized, once the latter are measured appropriately and the constraints that govern the mapping between them are properly formalized. We expect target uniformity to extend to additional languages, speaker populations, and feature–target pairings, but comprehensive understanding of the influence of each uniformity constraint will require considerable future research along each of these dimensions.

APPENDIX

Contrast weighting

For the isolated speech study, the categorical variables for the mixed-effects linear regression model were weighted effect coded with the following weights: place of articulation (*place*: +anterior = 1, –anterior = -1.19), voice (*voice*: voiceless = 1, voiced = -1.04), vowel height (*height*: high = 1, non-high = -0.42), vowel backness (*backness*: front = 1, non-front = -0.90), and gender (*gender*: female = 1, male = -2.11) (see Darlington 1990; te Grotenhuis et al. 2016). The dependent variable (mid-frequency peak) was centered at zero by subtracting the grand mean (μ = 28.08 ERB) from each value prior to analysis.

For the spontaneous speech study, the contrast weighting of the categorical variables was: place of articulation (*place*: +anterior = 1, –anterior = -4.12), voice (*voice*: voiceless = 1, voiced = -4.37), vowel height (*height*: high = 1, non-high = -0.53); vowel backness (*backness*: front = 1, non-front = -1.12), and gender (*gender*: female = 1, male = -0.80). The dependent variable (mid-frequency peak) was centered at zero by subtracting the grand mean (μ = 28.01 ERB) from each value.

TABLE 6. Range and median number of tokens per talker and sibilant fricative, and total number of tokens per sibilant in the American English isolated speech data. Two speakers did not produce any tokens of [ʒ] and were thus excluded in the counts of [ʒ] presented here.

Fricative	Range	Median	Total
s	16 – 25	24	527
z	15 – 25	24	520
ʃ	14 – 22	21	455
ʒ	10 – 25	23	424

TABLE 7. Descriptive statistics for each sibilant fricative in the American English isolated speech data. The mean and standard deviation were calculated from the population sample of talker-specific means. Ranges are reported for talker-specific means and standard deviations.

Measure	Fricative	Grand Mean	SD	Range of Talker Means	Range of Talker SDs
mid-frequency peak (ERB)	s	30.31	1.00	28.59 – 32.06	0.60 – 1.96
	z	30.41	1.02	28.61 – 31.93	0.73 – 2.87
	ʃ	25.40	1.39	22.87 – 27.20	0.48 – 2.71
	ʒ	25.22	1.31	22.96 – 27.64	0.64 – 2.97
mid-frequency peak (Hz)	s	5857	557	4844 – 6778	368 – 1071
	z	5858	567	4850 – 6705	440 – 1406
	ʃ	3431	544	2498 – 4181	181 – 1269
	ʒ	3354	508	2523 – 4426	188 – 1436
COG (Hz)	s	7946	1113	6158 – 10172	467 – 1404
	z	7244	912	5483 – 8621	546 – 2141
	ʃ	4305	652	3352 – 5210	244 – 1637
	ʒ	3979	583	2979 – 4997	239 – 1243

TABLE 8. Pearson correlation coefficients of talker means and corresponding SDs for each sibilant fricative in the American English isolated speech data.

Fricative	Mid-Frequency Peak (ERB)	Mid-Frequency Peak (Hz)	COG (Hz)
s	0.47	0.49	0.43
z	0.19	0.38	0.01
ʃ	0.54^{+}	0.72*	0.28
ʒ	0.36	0.54	0.27

* = $p < 0.001$, $^{+} = p < 0.01$

TABLE 9. Pearson correlation coefficients of talker means in the American English isolated speech data.

Measure	Fricative Pair	All	Female	Male
mid-frequency peak (ERB)	s – z	0.83*	0.80*	0.80
	ʃ – ʒ	0.91*	0.92*	0.74
	s – ʃ	0.60^{+}	0.50	0.41
	z – ʒ	0.32	0.34	-0.38
mid-frequency peak (Hz)	s – z	0.85*	0.82*	0.78
	ʃ – ʒ	0.88*	0.90*	0.58
	s – ʃ	0.61^{+}	0.52	0.46
	z – ʒ	0.30	0.29	-0.48
COG (Hz)	s – z	0.80*	0.73^{+}	0.99*
	ʃ – ʒ	0.79*	0.73^{+}	0.92*
	s – ʃ	0.56	0.55	0.09
	z – ʒ	0.33	0.31	0.27

* = $p < 0.001$, $^{+} = p < 0.01$

TABLE 10. Range and median number of tokens per talker and sibilant fricative, and total number of tokens per sibilant in the American English spontaneous speech from the Buckeye Corpus.

Fricative	Range	Median	Total
s	181 – 736	400	15,547
z	34 – 299	86	4,103
ʃ	48 – 217	111	4,320
ʒ	1 – 31	10.5	448

TABLE 11. Descriptive statistics for each sibilant fricative in the American English spontaneous speech from the Buckeye Corpus. The mean and standard deviation were calculated from the population sample of talker-specific means. Ranges are reported for talker-specific means and standard deviations. One speaker produced only one instance of [ʒ] and was thus excluded in the range of talker-specific [ʒ] SDs presented here.

Measure	Fricative	Grand Mean	SD	Range of Talker Means	Range of Talker SDs
mid-frequency peak (ERB)	s	28.80	1.40	23.95 – 30.31	0.37 – 3.26
	z	28.45	1.51	23.04 – 30.29	0.27 – 3.75
	ʃ	25.39	1.32	22.99 – 28.05	0.77 – 2.27
	ʒ	24.94	1.49	22.51 – 28.46	0.38 – 2.65
mid-frequency peak (Hz)	s	5034	706	2845 – 5838	249 – 1399
	z	4882	734	2561 – 5820	187 – 1612
	ʃ	3424	524	2613 – 4594	273 – 1018
	ʒ	3256	599	2399 – 4797	121.5 – 1163
COG (Hz)	s	5250	738	3336 – 6425	360 – 899
	z	4737	833	2771 – 6502	503 – 1587
	ʃ	3874	555	3013 – 4873	283 – 698
	ʒ	3444	606	2242 – 4760	100 – 1199

TABLE 12. Pearson correlation coefficients of talker means and corresponding SDs for each sibilant fricative in the American English spontaneous speech from the Buckeye Corpus.

Fricative	Mid-Frequency Peak (ERB)	Mid-Frequency Peak (Hz)	COG (Hz)
s	-0.23	-0.06	-0.12
z	-0.15	0.01	-0.16
ʃ	-0.18	0.25	0.19
ʒ	0.40	0.64*	0.35

* = $p < 0.001$, $^{+} = p < 0.01$

TABLE 13. Pearson correlation coefficients of talker means in the American English spontaneous speech from the Buckeye Corpus.

Measure	Fricative Pair	All	Female	Male
mid-frequency peak (ERB)	s – z	0.98*	0.91*	0.99*
	ʃ – ʒ	0.88*	0.77*	0.75*
	s – ʃ	0.73*	0.29	0.71*
	z – ʒ	0.60*	0.18	0.57
mid-frequency peak (Hz)	s – z	0.98*	0.90*	0.99*
	ʃ – ʒ	0.88*	0.77*	0.75*
	s – ʃ	0.73*	0.37	0.73*
	z – ʒ	0.62*	0.25	0.58^{+}
COG (Hz)	s – z	0.91*	0.76*	0.96*
	ʃ – ʒ	0.85*	0.70*	0.85*
	s – ʃ	0.78*	0.28	0.74*
	z – ʒ	0.58*	0.08	0.67^{+}

* = $p < 0.001$, $^{+} = p < 0.01$

REFERENCES

BEDDOR, PATRICE SPEETER; JAMES D. HARNSBERGER; and STEPHANIE LINDEMANN. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. ***Journal of Phonetics*** 30.591–627.

BLACKLOCK, OLIVER. 2004. ***Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods***. Southampton, UK: University of Southampton dissertation.

BROWMAN, CATHERINE P., and LOUIS M. GOLDSTEIN. 1989. Articulatory gestures as phonological units. ***Phonology*** 6.201–51.

BÜRKNER, PAUL-CHRISTIAN. 2017. brms: An R package for Bayesian multilevel models using Stan. ***Journal of Statistical Software*** 80.1–28.

BÜRKNER, PAUL-CHRISTIAN. 2018. Advanced Bayesian multilevel modeling with the R package brms. ***The R Journal*** 10.395–411.

BYRD, DANI, and ELLIOT L. SALTZMAN. 1998. Intragestural dynamics of multiple prosodic boundaries. ***Journal of Phonetics*** 26.173–99.

CARIGNAN, CHRISTOPHER. 2013. ***When nasal is more than nasal: The oral articulation of nasal vowels in two dialects of French***. Champaign-Urbana, IL: UIUC dissertation.

CARIGNAN, CHRISTOPHER; RYAN SHOSTED; CHILIN SHIH; and PANYING RONG. 2011. Compensatory articulation in American English nasalized vowels. ***Journal of Phonetics*** 39.668–82.

CHO, TAEHONG, and PETER LADEFOGED. 1999. Variation and universals in VOT: Evidence from 18 languages. ***Journal of Phonetics*** 27. 207–29.

CHODROFF, ELEANOR, and MELISSA M. BAESE-BERK. 2019. Constraints on variability in the voice onset time of L2 English stop consonants. ***Proceedings of the 19th International Congress of Phonetic Sciences***, Melbourne, 661–65.

CHODROFF, ELEANOR; ALESSANDRA GOLDEN; and COLIN WILSON. 2019. Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization. ***The Journal of the Acoustical Society of America*** 145.EL109–15.

CHODROFF, ELEANOR, and COLIN WILSON. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. ***Journal of Phonetics*** 61.30–47.

CHODROFF, ELEANOR, and COLIN WILSON. 2018. Predictability of stop consonant phonetics

across talkers: Between-category and within-category dependencies among cues for place and voice. ***Linguistics Vanguard*** 4.

CHODROFF, ELEANOR, and COLIN WILSON. 2020. Acoustic–phonetic and auditory mechanisms of adaptation in the perception of sibilant fricatives. ***Attention, Perception, and Psychophysics*** 82.2027–48.

CHOMSKY, NOAM, and MORRIS HALLE. 1968. ***The sound patterns of English***. New York: Harper.

CLEMENTS, G. N. 1985. The geometry of phonological features. ***Phonology*** 2.225–52.

CLEMENTS, G. N. 2003. Feature economy as a phonological universal. ***Proceedings of the 15th International Congress of Phonetic Sciences***, Barcelona, 371–74. Online: http://nickclements.free.fr/publications/2003e.PDF

COHN, ABIGAIL C. 1993. Nasalisation in English: Phonology or phonetics. ***Phonology*** 10.43–81.

COHN, ABIGAIL C., and MARIE K. HUFFMAN. 2014. ***Interface between phonology and phonetics***. Oxford University Press.

COOPER, ANDRÉ M. 1991. Laryngeal and oral gestures in English /p, t, k/. ***Proceedings of the 12th International Congress of Phonetic Sciences***, Aix-en-Provence, 50–54.

DARLINGTON, RICHARD. B. 1990. ***Regression and linear models***, ed. by J. D. Anker and B. Boylan. New York: McGraw-Hill Publishing Company.

DAVIDSON, LISA. 2016. Variability in the implementation of voicing in American English obstruents. ***Journal of Phonetics*** 54.35–50.

DEDIU, DAN; RICK JANSSEN; and SCOTT R. MOISIK. 2019. Weak biases emerging from vocal tract anatomy shape the repeated transmission of vowels. ***Nature Human Behaviour*** 3.1107–15.

DOCHERTY, GERARD. 1992. ***The timing of voicing in British English obstruents***. Berlin: Walter de Gruyter.

ECKERT, PENELOPE. 2008. Variation and the indexical field. ***Journal of Sociolinguistics*** 12.453–76.

EVANS, JAMES. 1996. ***Straightforward statistics for the behavioral sciences***. Pacific Grove, CA: Brooks/Cole Publishing.

EVERS, VINCENT; HENNING REETZ; and ADITI LAHIRI. 1998. Crosslinguistic acoustic categorization of sibilants independent of phonological status. ***Journal of Phonetics***

26.345–70.

FAYTAK, MATTHEW. 2018. ***Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Suzhou Chinese***. Berkeley, CA: UC Berkeley dissertation.

FLEMMING, EDWARD S. 2004. Contrast and perceptual distinctiveness. ***Phonetically based phonology***, ed. by Bruce Hayes, Ruth Kirchner, and Donca Steriade, 232–76. Cambridge: Cambridge University Press.

FLIPSEN, PETER; LAWRENCE SHRIBERG; GARY WEISMER; HEATHER KARLSSON; and JANE MCSWEENY. 1999. Acoustic characteristics of /s/ in adolescents. ***Journal of Speech, Language, and Hearing Research*** 42.663–77.

FORREST, KAREN; GARY WEISMER; PAUL MILENKOVIC; and RONALD N. DOUGALL. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. ***The Journal of the Acoustical Society of America*** 84.115–23.

FOULKES, PAUL; GERARD DOCHERTY; and DOMINIC WATT. 2001. On the emergence of structured phonological variation. ***University of Pennsylvania Working Papers in Linguistics*** 7.67–84.

FRIDLAND, VALERIE. 2001. The social dimension of the Southern Vowel Shift: Gender, age and class. ***Journal of Sociolinguistics*** 5.233–53.

FRUEHWALD, JOSEF. 2013. ***The phonological influence on phonetic change***. Philadelphia: University of Pennsylvania dissertation.

FRUEHWALD, JOSEF. 2017. The role of phonology in phonetic change. ***Annual Review of Linguistics*** 3.25–42.

FRUEHWALD, JOSEF. 2019. Is phonetic target uniformity phonologically, or sociolinguistically grounded? ***Proceedings of the 19th International Congress of Phonetic Sciences***, Melbourne, 681–85.

FUCHS, SUSANNE, and MARTINE TODA. 2010. Do differences in male versus female /s/ reflect biological or sociophonetic factors? ***Turbulent sounds: An interdisciplinary guide***, ed. by Susanne Fuchs, Martine Toda, and Marzena Żygis, 281–302.

GHOSH, SATRAJIT S.; MELANIE L. MATTHIES; EDWIN MAAS; ALEXANDRA HANSON; MARK TIEDE; LUCIE MÉNARD; FRANK H. GUENTHER; HARLAN LANE; and JOSEPH S. PERKELL. 2010. An investigation of the relation between sibilant production and somatosensory and auditory acuity. ***The Journal of the Acoustical Society of America***. 128.3079–87.

GLASBERG, BRIAN R.; and BRIAN C. J. MOORE. 1990. Derivation of auditory filter shapes from notched-noise data. ***Hearing Research*** 47.103–38.

GORDON, MATTHEW; PAUL BARTHMAIER; and KATHY SANDS. 2002. A cross-linguistic acoustic study of voiceless fricatives. ***Journal of the International Phonetic Association*** 32.141–74.

GUENTHER, FRANK H. 1994. A neural network model of speech acquisition and motor equivalent speech production. ***Biological Cybernetics*** 72.43–53.

GUENTHER, FRANK H. 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. ***Psychological Review*** 102.594–621.

GUY, GREGORY R. 2013. The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? ***Journal of Pragmatics*** 52.63–71.

GUY, GREGORY R., and FRANS HINSKENS. 2016. Linguistic coherence: Systems, repertoires and speech communities. ***Lingua*** 172–173.1–9.

HADDICAN, BILL; PAUL FOULKES; VINCENT HUGHES; and HAZEL RICHARDS. 2013. Interaction of social and linguistic constraints on two vowel changes in northern England. ***Language Variation and Change*** 25.371–403.

HALEY, KATARINA L.; RALPH N. OHDE; and ROBERT T. WERTZ. 2000. Precision of fricative production in aphasia and apraxia of speech: A perceptual and acoustic study. ***Aphasiology*** 14.619–34.

HALEY, KATARINA L.; ELIZABETH SEELINGER; KERRY CALLAHAN MANDULAK; and DAVID J. ZAJAC. 2010. Evaluating the spectral distinction between sibilant fricatives through a speaker-centered approach. ***Journal of Phonetics*** 38.548–54.

HALLE, MORRIS. 1983. On distinctive features and their articulatory implementation. ***Natural Language and Linguistic Theory*** 1.91–105.

HALLE, MORRIS. 1992. Phonological features. ***International Encyclopedia of Linguistics*** 3.207–12.

HAMANN, SILKE. 2010. The phonetics–phonology interface. ***Bloomsbury companion to phonology***, ed. by Nancy C. Kula, Bert Botma, and Kuniya Nasukawa, 202–24. New York: Bloomsbury Academic.

HEFFERNAN, KEVIN. 2004. Evidence from HNR that /s/ is a social marker of gender. ***Toronto Working Papers in Linguistics*** 23.71–84.

HOFFMAN, MATTHEW D., and ANDREW GELMAN. 2014. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. ***Journal of Machine Learning Research*** 15.1593–1623.

HOOLE, PHILIP, and MARIANNE POUPLIER. 2015. Interarticulatory Coordination. ***The handbook of speech production***, ed. by Melissa A. Redford, 133–57. Somerset, MA: Wiley.

HULLEBUS, MARC A.; STEPHEN J. TOBIN; and ADAMANTIOS I. GAFOS. 2018. Speaker-specific structure in German voiceless stop voice onset times. ***Proceedings of Interspeech 2018***, Hyderabad, 1403–07.

JEFFREYS, HAROLD. 1961. ***Theory of Probability***. Oxford: Oxford University Press.

JOHNSON, KEITH, and JACK B. MARTIN. 2001. Acoustic vowel reduction in Creek: Effects of distinctive length and position in the word. ***Phonetica*** 58.81–102.

JONGMAN, ALLARD; RATREE WAYLAND; and SERENA WONG. 2000. Acoustic characteristics of English fricatives. ***The Journal of the Acoustical Society of America*** 108.1252–63.

JOOS, MARTIN. 1948. Acoustic phonetics. ***Language*** 24.5–136.

KARY, ARTHUR; ROBERT TAYLOR; and CHRIS DONKIN. 2016. Using Bayes factors to test the predictions of models: A case study in visual working memory. ***Journal of Mathematical Psychology*** 72.210–19.

KEATING, PATRICIA A. 1985. Universal phonetics and the organization of grammars. ***Phonetic linguistics: Essays in honor of Peter Ladefoged***, ed. by Victoria. A. Fromkin, 115–32. London: Academic Press.

KEATING, PATRICIA A. 1988. The phonology-phonetics interface. ***Linguistics: The Cambridge survey,*** Volume I: Grammatical Theory, ed. by Frederick J. Newmeyer, 281–302. Cambridge: Cambridge University Press.

KEATING, PATRICIA A. 1990. Phonetic representations in a generative grammar. ***Journal of Phonetics*** 18.321–34.

KEATING, PATRICIA A. 2003. Phonetic and other influences on voicing contrasts. ***Proceedings of the 15th International Congress of Phonetic Sciences***, Barcelona, 20–23.

KINGSTON, JOHN. 2007. The phonetics-phonology interface. ***The Cambridge handbook of phonology***, ed. by Paul de Lacy, 401–34. Cambridge: Cambridge University Press.

KLEINSCHMIDT, DAVE F., and TIM F. JAEGER. 2015. Robust speech perception: Recognizing the familiar, generalizing to the similar, and adapting to the novel. ***Psychological Review***

122.148–203.

KOENIG, LAURA L.; CHRISTINE H. SHADLE; JONATHAN L. PRESTON; and CHRISTINE R. MOOSHAMMER. 2013. Toward improved spectral measures of /s/: Results from adolescents. ***Journal of Speech, Language, and Hearing Research*** 56.1175–89.

KRALJIC, TANYA, and ARTHUR G. SAMUEL. 2006. Generalization in perceptual learning for speech. ***Psychonomic Bulletin and Review*** 13.262–68.

LABOV, WILLIAM. 1972. ***Sociolinguistic patterns***. Philadelpha: University of Pennsylvania Press.

LABOV, WILLIAM. 2014. The sociophonetic orientation of the language learner. ***Advances in Sociophonetics*** 15.17–29.

LABOV, WILLIAM; INGRID ROSENFELDER; and JOSEF FRUEHWALD. 2013. One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis. ***Language*** 89.30–65.

LADD, D. ROBERT. 2014. Phonetics in phonology. ***Simultaneous structure in phonology***, 29–56. Oxford: Oxford University Press.

LAMETTI, DANIEL. R.; SAZZAD M. NASIR; and DAVID J. OSTRY. 2012. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. ***Journal of Neuroscience*** 32.9351–58.

LEHISTE, ILSE. 1970. ***Suprasegmentals***. Cambridge, MA: MIT Press.

LEVON, EREZ, and SOPHIE HOLMES-ELLIOTT. 2013. East End boys and West End girls: /s/-fronting in Southeast England. ***University of Pennsylvania Working Papers in Linguistics*** 19.111–20.

LI, FANGFANG; JAN R. EDWARDS; and MARY E. BECKMAN. 2007. Spectral measures for sibilant fricatives of English, Japanese, and Mandarin Chinese. ***Proceedings of the 16th International Congress of Phonetic Sciences***, Saarbrücken, 917–20.

LIBERMAN, ALVIN M.; FRANKLIN S. COOPER; DONALD P. SHANKWEILER; and MICHAEL STUDDERT-KENNEDY. 1967. Perception of the speech code. ***Psychological Review*** 74.431–61.

LILJENCRANTS, JOHAN; and BJÖRN LINDBLOM. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. ***Language*** 48.839–62.

LINDBLOM, BJÖRN. 1967. Vowel duration and a model of lip-mandible coordination. ***Speech, Music and Hearing: Quarterly Progress and Status Report*** 8.1–29.

LINDBLOM, BJÖRN. 1983. Economy of speech gestures. ***The production of speech***, ed. by Peter F. MacNeilage, 217–245. New York: Springer.

LINDBLOM, BJÖRN. 1986. Phonetic universals in vowel systems. ***Experimental phonology***, ed. by John J. Ohala and Jeri J. Jaeger, 13–44. Orlando: Academic Press.

LINDBLOM, BJÖRN, and IAN MADDIESON. 1988. Phonetic universals in consonant systems. ***Language, speech, and mind***, ed. by Larry M. Hyman and Charles N. Li, 62–78. London: Routledge.

LINVILLE, SUE ELLEN. 1998. Acoustic correlates of perceived versus actual sexual orientation in men's speech. ***Folia Phoniatrica et Logopaedica*** 50.35–48.

LOBANOV, BORIS M. 1971. Classification of Russian vowels spoken by different speakers. ***The Journal of the Acoustical Society of America*** 49.606–08.

LÖFQVIST, ANDERS, and HIROHIDE YOSHIOKA. 1984. Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. ***Speech Communication*** 3.279–89.

MADDIESON, IAN. 1995. Gestural economy. ***Proceedings of the 13th International Congress of Phonetic Sciences***, Stockholm, 574–578.

MADDIESON, IAN. 1997. Phonetic universals. ***Handbook of phonetic sciences***, ed. by William. J. Hardcastle and John Laver, 619–39. Oxford: Blackwell.

MANIWA, KAZUMI; ALLARD JONGMAN; and TRAVIS WADE. 2009. Acoustic characteristics of clearly spoken English fricatives. ***The Journal of the Acoustical Society of America*** 125.3962–73.

MASSARO, DOMINIC W. 1975. Preperceptual images, processing time, and perceptual units in speech perception. ***Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics***, ed. by Dominic W. Massaro, 25–50. London: Academic Press.

MAYE, JESSICA; RICHARD N. ASLIN; and MICHAEL K. TANENHAUS. 2008. The weckud wetch of the wast: Lexical adaptation to a novel accent. ***Cognitive Science*** 32.543–62.

MCMURRAY, BOB, and ALLARD JONGMAN. 2011. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. ***Psychological Review*** 118.219–46.

MÉNARD, LUCIE; JEAN-LUC SCHWARTZ; and JÉRÔME AUBIN. 2008. Invariance and variability in the production of the height feature in French vowels. ***Speech Communication*** 50.14–28.

MILLER, JOANNE L. 1994. On the internal structure of phonetic categories: A progress report. ***Cognition*** 50.271–85.

MOHR, BUURCKHARD. 1971. Intrinsic variations in the speech signal. ***Phonetica*** 23.65–93.

NAPOLI, DONNA JO; NATHAN SANDERS; and REBECCA WRIGHT. 2014. On the linguistic effects of articulatory ease, with a focus on sign languages. ***Language*** 90.424–56.

NARTEY, JONAS N.A. 1982. ***On fricative phones and phonemes***. Los Angeles: UCLA dissertation.

NEAREY, TERRANCE M. 1978. ***Phonetic feature systems for vowels***. Alberta: University of Alberta dissertation.

NEAREY, TERRANCE M., and PETER F. ASSMANN. 2007. Probabilistic "sliding template" models for indirect vowel normalization. In Solé, Beddor, & Ohala, 246–70.

NEWMAN, ROCHELLE S.; SHERYL A. CLOUSE; and JESSICA L. BURNHAM. 2001. The perceptual consequences of within-talker variability in fricative production. ***The Journal of the Acoustical Society of America*** 109.1181–96.

NICENBOIM, BRUNO; DANIEL SCHAD; and SHRAVAN VASISHTH. 2021. ***An introduction to Bayesian data analysis for cognitive science***, MS. Online: https://vasishth.github.io/bayescogsci/book/

NIEBUHR, OLIVER; MEGHAN A. CLAYARDS; CHRISTINE MEUNIER; and LEONARDO LANCIA. 2011. On place assimilation in sibilant sequences—Comparing French and English. ***Journal of Phonetics*** 39.429–51.

OHALA, JOHN J. 1972. How is pitch lowered? ***The Journal of the Acoustical Society of America*** 52.124.

OHALA, JOHN J. 1979. The contribution of acoustic phonetics to phonology. ***Frontiers of speech communication research*** , ed. by Björn Lindblom and S. Öhman, 355–63. London: Academic Press.

OHALA, JOHN J. 1980. Moderator's summary of symposium on "Phonetic universals in phonological systems and their explanation." ***Proceedings of the 9th International Congress of Phonetic Sciences***, Copenhagen, 181–94.

OHALA, JOHN J. 1990. There is no interface between phonology and phonetics: A personal view. ***Journal of Phonetics*** 18.153–71.

OHALA, JOHN J., and BRIAN W. EUKEL, B. 1987. Explaining the intrinsic pitch of vowels. ***In***

honor of Ilse Lehiste, ed. by Robbert Channon and Linda Schocky, 207–215. New York: De Gruyter Mouton.

OUSHIRO, LIVIA. 2019. Linguistic uniformity in the speech of Brazilian internal migrants in a dialect contact situation. ***Proceedings of the 19th International Congress of Phonetic Sciences***, Melbourne, 686–90.

PETERSON, GORDON E. and HAROLD L. BARNEY. 1952. Control methods used in a study of the vowels. ***The Journal of the Acoustical Society of America*** 24.175–84.

PIERREHUMBERT, JANET B. 1990. Phonological and phonetic representation. ***Journal of Phonetics*** 18.375–94.

PISONI, DAVID B. and JAMES R. SAWUSCH. 1975. Some stages of processing in speech perception. ***Structure and process in speech perception***, ed. by Antonie Cohen and Sibout G. Nooteboom, 16–35. Heidelberg: Springer.

PITT, MARK A.; KEITH JOHNSON; ELIZABETH HUME; SCOTT KIESLING; and WILLIAM RAYMOND. 2005. The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. ***Speech Communication*** 45.89–95.

PODESVA, ROBERT J., and 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*** 20.128–37.

REIDY, PATRICK F. 2015. A comparison of spectral estimation methods for the analysis of sibilant fricatives. ***The Journal of the Acoustical Society of America*** 137.EL248–54.

REIDY, PATRICK F. 2016. Spectral dynamics of sibilant fricatives are contrastive and language specific. ***The Journal of the Acoustical Society of America*** 140.2518–29.

SALESKY, ELIZABETH; ELEANOR CHODROFF; TIAGO PIMENTEL; MATTHEW WIESNER; RYAN COTTERELL; ALAN W. BLACK; and JASON EISNER. 2020. A corpus for large-scale phonetic typology. ***Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics***, Online, 4526–46.

SCHWARTZ, MARTIN F. 1968. Identification of speaker sex from isolated, voiceless fricatives. ***The Journal of the Acoustical Society of America*** 43.1178–79.

SCHWARTZ, JEAN-LUC; LOUIS-JEAN BOË; and CHRISTIAN ABRY. 2007. Linking the dispersion-focalization theory and the maximum utilization of the available distinctive features principle in a perception-for-action-control theory. In Solé, Beddor, & Ohala, 104–24.

SCHWARTZ, JEAN-LUC, and LUCIE MÉNARD. 2019. Structured idiosyncrasies in vowel systems, MS. Online: https://doi.org/10.31219/osf.io/b6rdv

SHADLE, CHRISTINE H.; WADE CHEN; and DOUGLAS H. WHALEN. 2016. Stability of the main resonance frequency of fricatives despite changes in the first spectral moment. ***The Journal of the Acoustical Society of America*** 140.3219–20.

SHADLE, CHRISTINE H.; LAURA L. KOENIG; and JONATHAN L. PRESTON. 2014. Acoustic characterization of /s/ spectra of adolescents: Moving beyond moments. ***Proceedings of Meetings on Acoustics*** 12.1–20.

SHADLE, CHRISTINE H., and SHEILA J. MAIR. 1996. Quantifying spectral characteristics of fricatives. ***Proceedings of the 4th International Conference on Spoken Language Processing***, Philadelphia, 1521–24.

SHAW, JASON A.; ADAMANTIOS I. GAFOS; PHILIP HOOLE; and CHAKIR ZEROUAL. 2009. Syllabification in Moroccan Arabic: Evidence from patterns of temporal stability. ***Phonology*** 26.187–215.

SHOSTED, RYAN; CHRISTOPHER CARIGNAN; and PANYING RONG. 2012. Managing the distinctiveness of phonemic nasal vowels: Articulatory evidence from Hindi. ***The Journal of the Acoustical Society of America*** 131.455–65.

SILBERT, NOAH, and KENNETH DE JONG. 2008. Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. ***The Journal of the Acoustical Society of America*** 123.2769–79.

SOLÉ, MARIA JOSEP. 1992. Phonetic and phonological processes: The case of nasalization. ***Language and Speech*** 35.29–43.

SOLÉ, MARIA JOSEP; PATRICIA SPEETER BEDDOR; and MANJARI OHALA (eds.). 2007. ***Experimental approaches to phonology***. Oxford: Oxford University Press.

SONDEREGGER, MORGAN; JANE STUART-SMITH; THEA KNOWLES; RACHEL MACDONALD; and TAMARA RATHCKE. 2020. Structured heterogeneity in Scottish stops over the 20th century. ***Language*** 96.94–125.

STRAND, ELIZABETH A. 1999. Uncovering the role of gender stereotypes in speech perception. ***Journal of Language and Social Psychology*** 18.86–100.

STRAND, ELIZABETH A., and KEITH JOHNSON. 1996. Gradient and visual speaker normalization in the perception of fricatives. ***Natural Language Processing and Speech Technology:***

Results of the 3rd KONVENS Conference, Bielefeld, 14–26.

STUART-SMITH, JANE; CLAIRE TIMMINS; and ALAN WRENCH. 2003. Sex and gender differences in Glaswegian /s/. ***Proceedings of the 15th International Congress of Phonetic Sciences***, Barcelona, 1851–54.

TANNER, JAMES; MORGAN SONDEREGGER; and JANE STUART-SMITH. 2019. Structured speaker variability in spontaneous Japanese stop contrast production. ***Proceedings of the 19th International Congress of Phonetic Sciences***, Melbourne, 666–70.

TE GROTENHUIS, MANFRED; BEN PELZER; ROB EISINGA; RENSE NIEUWENHUIS; ALEXANDER SCHMIDT-CATRAN; and RUBEN KONIG. 2016. When size matters: Advantages of weighted effect coding in observational studies. ***International Journal of Public Health*** 62.1–5.

THEODORE, RACHEL M., and JOANNE L. MILLER. 2010. Characteristics of listener sensitivity to talker-specific phonetic detail. ***The Journal of the Acoustical Society of America*** 128.2090–99.

THOMSON, DAVID J. 1982. Spectrum estimation and harmonic analysis. ***Proceedings of the IEEE*** 70.1055–96.

TURK, ALICE E., and STEFANIE SHATTUCK-HUFNAGEL. 2014. Timing in talking: What is it used for, and how is it controlled? ***Philosophical Transactions of the Royal Society B*** 369.

VANPAEMEL, WOLF. 2010. Prior sensitivity in theory testing: An apologia for the Bayes factor. ***Journal of Mathematical Psychology*** 54.491–98.

VASISHTH, SHRAVAN; BRUNO NICENBOIM; MARY E. BECKMAN; FANGFANG LI; and EUN JONG KONG. 2018. Bayesian data analysis in the phonetic sciences: A tutorial introduction. ***Journal of Phonetics*** 71.147–61.

VOLENEC, VENO, and CHARLES REISS. 2017. Cognitive phonetics: The transduction of distinctive features at the phonology-phonetics interface. ***Biolinguistics*** 11.251–94.

WATT, DOMINIC. 2000. Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? ***Language Variation and Change*** 12.69–101.

WESTBURY, JOHN R., and PATRICIA A. KEATING. 1980. Central representation of vowel duration. ***The Journal of the Acoustical Society of America*** 67.S37.

WHALEN, DOUGLAS H., and ANDREA G. LEVITT. 1995. The universality of intrinsic f0 of vowels. ***Journal of Phonetics*** 23.349–66.

WILSON, COLIN, and ELEANOR CHODROFF. 2017. Uniformity of inherent vowel duration across

speakers of American English. ***The Journal of the Acoustical Society of America*** 142.2580.

YAO, YAO. 2009. Understanding VOT variation in spontaneous speech. ***UC Berkeley PhonLab Annual Report*** 5.29–43.

YU, ALAN C. L. 2019. On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. ***Laboratory Phonology*** 10.1–29.

YUAN, JIAHONG, and MARK Y. LIBERMAN. 2008. Speaker identification on the SCOTUS corpus. ***Proceedings of Acoustics '08***, Paris, 5687–90.

YUNUSOVA, YANA; KRISTY RUDY; and MELANIE BALJKO. 2012. Positional targets for lingual consonants defined using electromagnetic articulography. ***The Journal of the Acoustical Society of America*** 132.1027–38.

ZIMMAN, LAL. 2017. Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/. ***Language in Society*** 46.339–70.

ZSIGA, ELIZABETH. 1997. Features, gestures, and Igbo vowels: An approach to the Phonology-Phonetics Interface. ***Language*** 73.227–74.