HOW VARIABLE ARE ENGLISH SIBILANTS?

Morgan Sonderegger[*], Jane Stuart-Smith[†], Jeff Mielke[‡], The SPADE Consortium

McGill University[*]; University of Glasgow[†]; North Carolina State University[‡]
morgan.sonderegger@mcgill.ca, Jane.Stuart-Smith@glasgow.ac.uk, jimielke@ncsu.edu

ABSTRACT

This paper takes a corpus-phonetic approach to consider variability in word-initial, prevocalic English sibilants in 5k speakers, 235k tokens, from 27 geo-social-ethnic regions of North America and the British Isles. We analyse ERB-transformed, spectral peak measures calculated from amplitude-normalised multitaper spectra using a 'distributional' Bayesian mixed-effects regression which explicitly models token, speaker, and region-level variability. Following previous phonetic and sociolinguistic research we expected English /s/ to be more variable than /ʃ/. The results, however, differ according to the level at which we consider variability. Across English regions, /s/ and /ʃ/ show a similar degree of variability. Across speakers within-region, /s/ is generally more variable than /ʃ/, and within speakers, /s/ is generally more variable by token than /ʃ/, both results likely reflecting linguistic and social-indexical sources of variation—such as gender, which has a greater effect on spectral peak for /s/ than for /ʃ/.[1]

Keywords: sibilants, corpus phonetics, variability, sociophonetics, spectral analysis

1. INTRODUCTION

The sibilant fricatives /s ʃ/, e.g. *seat, sheet*, contrast in North American and British Isles Englishes, with /s/ usually alveolar/denti-alveolar, and /ʃ/ palatoalveolar, also with lip-rounding. Bar subtle known differences in sibilant production from e.g. possible anatomical differences in vocal cavity size [1], English sibilants are assumed to be produced in fairly similar ways and to be acoustically stable across English dialects [2].

At the same time, experimental studies have shown that /s/ is more subject to coarticulatory pressures than /ʃ/, including lip-rounding [3] and articulatory and auditory retraction [4, 5]. This is also reflected in diachrony: /s/ is involved in contextually-induced sound change more often than /ʃ/ [6]. And numerous sociophonetic studies [7] have demonstrated the performative nature of /s/ productions linked to local social-indexical meanings, overriding anatomical differences [8], whilst /ʃ/ is generally assumed to be relatively less socially informative [9]. However, the very large body of phonetic research on English sibilants [2, 10, 11, 12] shows a dialect bias towards speech largely from younger educated adults speaking standard varieties of American and UK English [13]. Sociophonetic research also shows an observational bias towards /s/, with much less attention paid to /ʃ/. Here we take a corpus-phonetic approach, increasing sample size and dialect diversity, to ask: *Is /s/ more variable than /ʃ/ in English?*

2. METHOD

This study presents results from 32 public and private spoken corpora, phonetic and (socio)linguistic, comprising spontaneous and read speech, from the British Isles (England, Scotland, Wales, Northern Ireland and the Republic of Ireland), and from America and Canada. The 5042 speakers (2544 female and 2498 male) were subsequently grouped into 27 broad geo-social-ethnic dialect regions (here 'dialects'), following e.g. [14, 15].[2] All dialects contain spontaneous speech, and some also contain read speech. They may also have speakers of different ages, though we restricted our chronological window to recordings made from the 1990s to the 2010s.

We used [16] to import the force-aligned corpora, and then to extract and analyse all instances of word-initial, prevocalic /s ʃ/ in stressed syllables. We examined the empirical distributions of durations and excluded tokens with durations below 35ms (too short for spectral analysis) or above 400ms (likely alignment errors) (5.2k tokens/6% of the data). We also excluded 4.1k tokens (1.7% of the data) with unrealistically low spectral peak based on the empirical distribution for each corpus. The final dataset consists of 235,582 tokens from 27 regions, 5042 speakers, and 1429 words.

Measuring the main spectral peak in sibilants from female speakers in 16kHz recordings is problematic, as the peak is often above the Nyquist frequency (8kHz). We therefore analysed the main spectral peak only using recordings with sampling rates of 22kHz or higher, which entailed excluding corpora from the full set [17] sampled at 16kHz. We also excluded recordings of 255 individual speakers after visual inspection of sibilant spectra. Often these were recordings which appeared to be high-sampling-rate digitized versions of low-quality original recordings, lacking energy at frequencies where sibilant noise is expected.

We created multitaper spectra from the middle 25ms of each sibilant interval using the `spectRum` package [18] for `R`, with 8 tapers, a bandwidth of 4, and no preemphasis. To compensate for varying recording conditions, we applied a spectrum normalization scheme. We divided all corpora into speech ‘utterances’ separated by segments of 150ms+ of non-speech in the forced-alignment, then sampled all utterance and non-utterance intervals at 1000 time points. Some corpora with many short sound recordings had no non-utterance intervals, so we measured utterance intervals instead. Within a given corpus, amplitudes were converted to decibels using the maximum frequency bin value for the utterance intervals as the 0 dB reference point. Spectra were then scaled, for each frequency bin, such that 1 equals the 90th percentile amplitude of utterance intervals and 0 equals the 10th percentile amplitude of non-utterance intervals. The main spectral peak was then measured as the frequency of the maximum amplitude between 1-11kHz.

Acoustic studies on sibilants vary in whether they report results in Hz or non-linear scales like Bark or ERB, and the results, especially for sibilant variation, are not always consistent across the two scales [12, 2, 19]; sociophonetic studies tend to report Hz. Given our interest in variability within and across the phonological categories (c.f. e.g. [20]), we transformed the peak frequency to ERB.

Our goals for modeling spectral `peak` were to assess three ways in which /s/ could be more variable than /ʃ/: **(1)** across dialects, **(2)** across speakers, within a given dialect, and **(3)** across tokens, within individual speakers. We modeled `peak` using a Bayesian regression model, fitted in Stan/brms [21, 22] using weakly-informative regularizing priors [23], which consists of two linear mixed-effects models (LMM): one for the mean of `peak`, as used in any corpus-phonetic analysis of spectral peak (e.g. [12]), and one for the *variance* of `peak` (specified as $\log(\sigma)$)—the amount of by-token variability, which is typically assumed to be constant, but in this kind of ‘distributional regression’ model [21] can itself vary, specifically to differ between /s/ and /ʃ/ [24].

The first LMM contains fixed effects for Onset (/s/ vs. /ʃ/), log-transformed sibilant Duration (expected: /s/ < /ʃ/), and their interaction, as well as speaker `gender` (expectation: F > M), and an interaction with `onset`, $\beta_{\text{gender:onset}}$, which partially captures comparison **(2)**, and allows for the F − M difference to differ between /s/ and /ʃ/. We include random effects to capture variability in `peak` for /s/ and /ʃ/ separately, by `dialect` and by `speaker` nested within `dialect`. These four random-effect variances, which we call $\sigma_{d,/s/}$, $\sigma_{d,/ʃ/}$, $\sigma_{s,/s/}$, and $\sigma_{s,/ʃ/}$, address **(1)** and **(2)** for an ‘average speaker’ and an ‘average dialect’, respectively. We also include near-maximal random slopes for `dialect` and `speaker`, as well as a by-`word` random intercept, which serves as a rough control for linguistic factors (e.g. following vowel height) beyond `onset` which could affect `peak`. The second LMM includes a fixed effect term of `onset` ($\beta_{\sigma,\text{onset}}$) which allows σ to differ between /s/ and /ʃ/, addressing **(3)**. Maximal by-`dialect` and by-`speaker` random effects are also included to accurately estimate this term.

3. RESULTS

We primarily report results relevant for our research questions **(1-3)**. We do so by summarizing the posterior distribution of a quantity of interest computed from model coefficients—e.g. the difference between $\sigma_{d,/s/}$ and $\sigma_{d,/ʃ/}$ for **(1)**—using the median, 95% credible interval (CredI), and probability of effect direction p_d, which are roughly like summarizing an effect for a frequentist model (with p_d = 1 minus p-value); see e.g. [23].

3.1. /s/ vs. /ʃ/ variability across dialects

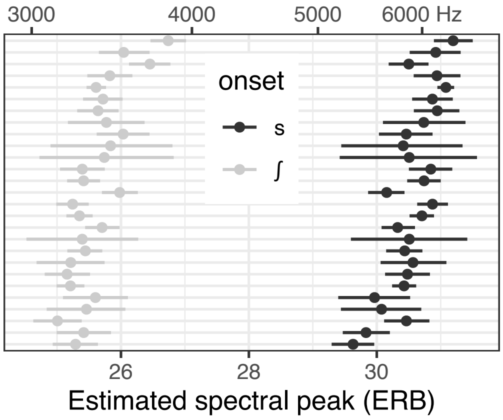

Figure 1: Estimated spectral peaks for English /s/ and /ʃ/ by dialect (one row per dialect). In all plots, dots/lines indicate posterior medians and 95% CredI’s.

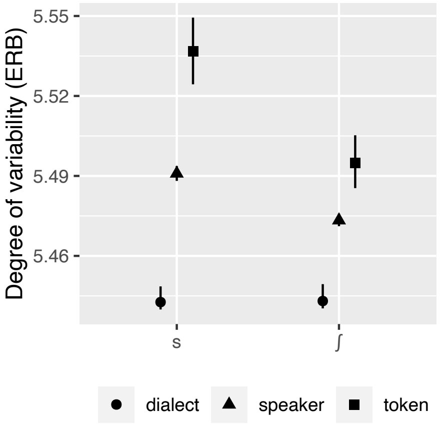

Figure 2: Estimated degree of variability for English /s/ and /ʃ/ by-dialect, by-speaker within-dialect ('speaker'), and by-token within-speaker.

Figure 1 shows spectral peak values estimated by the model for North American and British Isles English /s/ and /ʃ/. Note that dialects differ in certainty of estimates, corresponding to differing sample sizes, from $n < 250$ (3 rows with largest errorbars) to $n > 30$k (3 rows with smallest errorbars). Three findings are apparent. First, there is a gap between the highest /ʃ/ mean and the lowest /s/ mean, even accounting for uncertainty in model predictions. Second, there is substantial variability in the spectral peak location across dialects for both /s/ and /ʃ/. Third, the degree of variability does not clearly differ between /s/ and /ʃ/. This last point, which addresses **(1)**, can be shown by examining the estimated degree of by-dialect variability for each sibilant: the dots in Figure 2, $\sigma_{d,/s/}$ and $\sigma_{d,/ʃ/}$, are essentially the same, as confirmed by a hypothesis test ($\sigma_{d,/s/}$ - $\sigma_{d,/ʃ/}$: median = 0 ERB, 95% CredI = $[-0.005, 0.004]$, $p_d = 0.44$, BF = 0.77).

Two further observations emerge from Figure 1. Our finding that /s/ and /ʃ/ are equally variable cross-dialectally depends on using the ERB scale, closer to auditory processing, than to the linear Hz scale (top axis), where /s/ shows more dialectal variation than /ʃ/. Second, dialects with lower /s/ peaks seem to have lower /ʃ/ peaks. The model's estimated correlation is 0.29 (95% CredI = $[-0.03, 0.57]$, $p_d =$ 0.93, BF = 14.0), which offers tentative support for 'contrast uniformity' of the sibilant place contrast, posited by [12], at the dialect level.

3.2. /s/ vs. /ʃ/ variability across speakers within dialects

Figure 3 shows the distribution of speaker /s/ and /ʃ/ peak as offsets from dialect means, i.e. how much every speaker differs in their realization of /s/ and /ʃ/ from what is expected given their dialect (see Figure 1). The widths of the distributions correspond to the estimated degrees of variability in sibilant spectral peak, $\sigma_{s,/s/}$ and $\sigma_{s,/ʃ/}$ (Figure 2: triangles). A hypothesis test addressing **(2)** confirms that the 'wider' distribution visible for /s/ (dark grey) is indeed more variable than /ʃ/ (light grey) ($\sigma_{s,/s/} - \sigma_{s,/ʃ/}$: median = 0.018 ERB, 95% CredI = $[0.015, 0.020]$, $p_d = 1$). We also find a difference in variability by speaker gender: the difference in spectral peak between male and female speakers is larger for /s/ than for /ʃ/ ($\beta_{\text{gender:onset}}$: 2.0 ERB, 95% CredI = $[1.9, 2.2]$, $p_d = 1$); see [25].

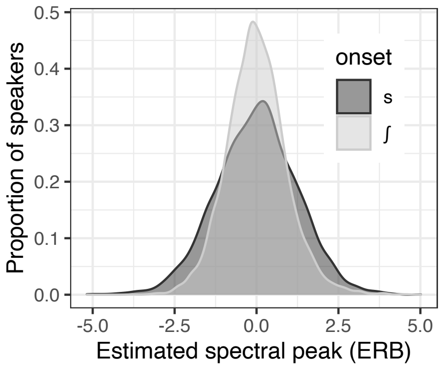

Figure 3: Distribution of estimated speaker offsets from dialect means (using posterior medians) for English /s/ and /ʃ/.

3.3. /s/ vs. /ʃ/ variability within speakers

Figure 2 (squares) addresses question **(3)**, showing the estimated amount of within-speaker variability for /s/ and /ʃ/, effectively capturing by-token variability for an 'average speaker' within an 'average dialect'. /s/ is generally more variable than /ʃ/, as confirmed by a hypothesis test ($\beta_{\sigma,\text{onset}}$: median = 0.08, 95% CredI = [0.04, 0.13], $p_d = 1$).

However, this result hides considerable variation across dialects and speakers. Figure 4 shows the estimated difference in within-speaker variability between /s/ and /ʃ/ for an 'average speaker' for *each* dialect (omitting the 3 dialects with $n < 250$). While most (16/27) dialects show greater variability for /s/ (95% CredI entirely positive), there is no clear difference in variability between /s/ and /ʃ/ in five, and in three, /s/ is *less* variable than /ʃ/ (95% CredI entirely negative). The five 'unclear' dialects may be inconclusive without more data; however, we can be sure that /s/ is not uniformly more variable than /ʃ/, at the token level in *all* dialects. As for variation within speakers, the actual predicted /s/-/ʃ/

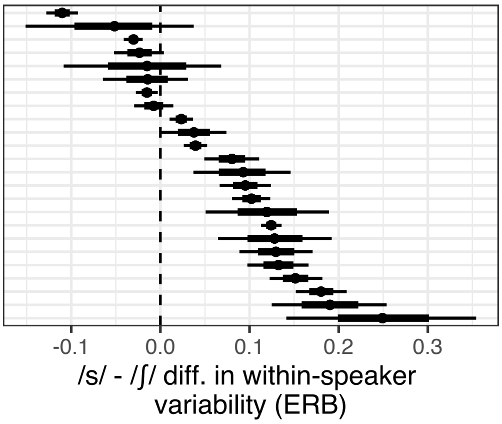

Figure 4: Estimated within-speaker variability between English /s/ and /ʃ/ by dialect.

variability differences are less informative because most speakers have sample size too small to say whether /s/ or /ʃ/ is more variable. Instead, we can examine the model's prediction for 'what proportion of speakers have /s/ variability > /ʃ/ variability?'. This fraction is 0.66 (95% CredI: [0.65, 0.68]), suggesting that /s/ is more variable at the token level for a majority of speakers, but not all.

4. DISCUSSION

This study considers acoustic invariance in English /s ʃ/ on a large scale. We find a robust psychoacoustic contrast between the sibilants across dialects, as consistently observed from previous lab studies [2, 20]. Whatever the precise articulations used across English, this sharp overall difference resonates with [26]'s assumption that the English phonological contrast exploits the abrupt acoustic consequences of a gradient shift in articulatory place and tongue shape, enhanced by lip-rounding.

Our main question is whether /s/ is more variable than /ʃ/ [7, 9], in a nationally, regionally, socially, and ethnically diverse sample of English dialect speakers. At the level of dialect, /s/ and /ʃ/ are equally variable. In contrast, /s/ is more variable than /ʃ/ across speakers within dialect, including gender differences being larger for /s/ [25]. Within speakers at the token level, /s/ is more variable than /ʃ/ on average, but not consistently across all dialects and speakers. At all levels we find greater variability in /ʃ/ than anticipated (c.f. [5]).

Thus, while we find that /s/ is 'more variable', the answer is more nuanced than expected. A major reason for this is our analytical choice of a non-linear, auditory scale (ERB, but we expect e.g. Bark would look similar) over Hz. In Hz, /s/ is more variable than /ʃ/ across dialects (Figure 1), and re-running our analysis in Hz also shows greater variability in /s/ across the board (as shown in the OSF project^1). This is partly because differences at higher frequencies are exaggerated in Hz, an observational bias which is undone by using a non-linear scale, closer to auditory processing. Discrepancies between Hz and auditorily-scaled measures have been noted before [2, 12], but these discrepancies are especially large for research questions about *variance*, the focus here, rather than *means*, as in most previous work on sibilants. This has implications for sociophonetic research, which often uses Hz measures to infer social-indexical meanings, especially for gendered, social and ethnic identities e.g. [27, 28]. The broader point is that research questions about *variability* in any frequency measure will be greatly affected by the choice of scale, especially for higher-frequency (e.g. sibilant energy, stop bursts) versus lower-frequency spectral measures (e.g. vowel formants).

What are the sources of variation for greater across- and within-speaker variability in /s/? There are likely several factors, including greater susceptibility to coarticulatory, physiological and performative social-indexical variation [1, 7]. To this we can add additional factors in our sample, including speaker age, speech style, and further structured sociolectal variation which may be subsumed in the broad dialect groupings used here.

Finally, the relatively high variability in spectral peak for /ʃ/ compared to that of /s/ across dialects and speakers seems surprising. But perhaps our expectations about what English sibilants 'should' be like are skewed by theoretical attention on /s/ [29], and/or observational—and so theoretical—biases, unwittingly informed by previous smaller-scale studies on less heterogeneous dialects. Further corpus-phonetic work on these, and other sounds will help us to appreciate better the scope of variability in English sounds (cf. [30, 31, 32]).

5. REFERENCES

[1] S. Fuchs and M. Toda, "Do differences in male versus female /s/ reflect biological or sociophonetic factors?" in *Turbulent Sounds*, S. Fuchs, M. Toda, and M. Zygis, Eds. Berlin: de Gruyter, 2010, pp. 281–302.

[2] V. Evers, H. Reetz, and A. Lahiri, "Crosslinguistic acoustic categorization of sibilants independent of phonological status," *J. Phonetics*, vol. 26, no. 4, pp. 345–370, 1998.

[3] L. L. Koenig, C. H. Shadle, J. L. Preston, and C. R. Mooshammer, "Toward improved spectral measures of /s/: Results from adolescents,"

J. Sp. Lang. Hear. Res., vol. 56, no. 4, pp. 1175–1189, 2013.

[4] A. Baker, D. Archangeli, and J. Mielke, "Variability in American English s-retraction suggests a solution to the actuation problem," *Language Variation and Change*, vol. 23, no. 3, pp. 347–374, 2011.

[5] D. Recasens and C. Rodríguez, "A study on coarticulatory resistance and aggressiveness for front lingual consonants and vowels using ultrasound," *J. Phonetics*, vol. 59, pp. 58–75, 2016.

[6] M. Stevens and J. Harrington, "The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English," *J. Phonetics*, vol. 58, pp. 118–134, 2016.

[7] E. Levon, M. Maegaard, and N. Pharao, Eds., "The sociophonetics of /s/," *Linguistics*, vol. 55, no. 5, 2017.

[8] J. Stuart-Smith, "Empirical evidence for gendered speech production: /s/ in Glaswegian," in *Change in Phonology: Papers in Laboratory Phonology 9*, J. Cole and J. Hualde, Eds. Berlin: de Gruyter, 2007, pp. 65–86.

[9] P. Eckert and S. McConnell-Ginet, *Language and Gender*, 2nd ed. Cambridge: Cambridge University Press, 2013.

[10] P. Flipsen Jr., L. Shriberg, G. Weismer, H. Karlsson, and J. McSweeny, "Acoustic characteristics of /s/ in adolescents," *J. Sp. Lang. Hear. Res.*, vol. 42, no. 3, pp. 663–677, 1999.

[11] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.*, vol. 108, pp. 1252–1263, 2000.

[12] E. Chodroff and C. Wilson, "Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English," *Language*, vol. 98, no. 2, 2022.

[13] J. M. Scobbie and J. Stuart-Smith, "Socially-stratified sampling in laboratory-based phonological experimentation," in *The Oxford Handbook of Laboratory Phonology*, A. C. Cohn, C. Fougeron, and M. K. Huffman, Eds. Oxford: Oxford University Press, 2012, pp. 607–621.

[14] A. Hughes, P. Trudgill, and D. Watt, *English Accents and Dialects*, 5th ed. London/New York: Routledge, 2013.

[15] W. Labov, S. Ash, and C. Boberg, *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: de Gruyter, 2006.

[16] M. McAuliffe, A. Coles, M. Goodale, S. Mihuc, M. Wagner, J. Stuart-Smith, and M. Sonderegger, "ISCAN: A system for integrated phonetic analyses across speech corpora," in *Proc. 19th ICPhS*, Melbourne, 2019, pp. 1322–1326.

[17] "SPeech Across Dialects of English," 2018. [Online]. Available: spade.glasgow.ac.uk

[18] P. F. Reidy, "spectRum," 2013, R package.

[19] ——, "Spectral dynamics of sibilant fricatives are contrastive and language specific," *J. Acoust. Soc. Am.*, vol. 140, pp. 2518–2529, 2016.

[20] P. Boersma and S. Hamann, "The evolution of auditory dispersion in bidirectional constraint grammars," *Phonology*, vol. 25, no. 2, pp. 217–270, 2008.

[21] P.-C. Bürkner, "Advanced Bayesian multilevel modeling with the R package brms," *The R Journal*, vol. 10, no. 1, pp. 395–411, 2018.

[22] Stan Development Team *et al.*, "Stan modeling language user's guide and reference manual," 2019, version 2.29. [Online]. Available: https://mc-stan.org

[23] B. Nicenboim, D. J. Schad, and S. Vasishth. (2021) An introduction to Bayesian data analysis for cognitive science. [Online]. Available: https://vasishth.github.io/bayescogsci/book/

[24] L. A. Ciaccio and J. Veríssimo, "Investigating variability in morphological processing with Bayesian distributional models," *Psych. Bull. Rev.*, vol. 29, no. 6, pp. 2264–2274, 2022.

[25] J. J. Holliday, P. F. Reidy, M. E. Beckman, and J. Edwards, "Quantifying the robustness of the English sibilant fricative contrast in children," *J. Sp. Lang. Hear. Res.*, vol. 58, pp. 622–637, 2015.

[26] K. N. Stevens, "On the quantal nature of speech," *J. Phonetics*, vol. 17, no. 1, pp. 3–45, 1989.

[27] R. J. Podesva and S. Kajino, "Sociophonetics, gender, and sexuality," in *The Handbook of Language, Gender, and Sexuality*, S. Ehrlich, M. Meyerhoff, and J. Holmes, Eds. Wiley Online Library, 2014, pp. 103–122.

[28] N. Pharao, K. L. Appel, V. Wolter, and J. Thøgersen, "Raising of /a/ in Copenhagen Danish – Perceptual consequences over two generations," in *Proc. 18th ICPhS*, Glasgow, 2015.

[29] J. Stuart-Smith, "Changing perspectives on /s/ and gender over time in Glasgow," *Linguistics Vanguard*, vol. 6, no. s1, 2020.

[30] M. Y. Liberman, "Corpus phonetics," *Annual Review of Linguistics*, vol. 5, pp. 91–107, 2019.

[31] E. Chodroff, A. Golden, and C. Wilson, "Covariation of stop voice onset time across languages: Evidence for a universal constraint on phonetic realization," *J. Acoust. Soc. Am.*, vol. 145, pp. EL109–EL115, 2019.

[32] J. Tanner, M. Sonderegger, J. Stuart-Smith, and J. Fruehwald, "Toward 'English' Phonetics: Variability in the Pre-consonantal Voicing Effect Across English Dialects and Speakers," *Frontiers in Artificial Intelligence*, vol. 3, 2020.

[1] We acknowledge funding from ESRC (ES/R003963/1) to Stuart-Smith; SSHRC (869-2016-0006, 435-2017-0925) and NSERC (RGPDD 501771-16) to Sonderegger; and NSF (SMA-1730479) to Mielke.

[2] Details of our data and analysis can be found in an OSF project at https://doi.org/10.17605/OSF.IO/XUBQM.