



5-29-2020

## Toward "English" Phonetics: Variability in the Pre-Consonantal Voicing Effect across English Dialects and Speakers

James Tanner  
*McGill University, Canada*

Morgan Sonderegger  
*McGill University, Canada*

Jane Stuart-Smith  
*University of Glasgow, UK*

Josef Fruehwald  
*University of Kentucky, jfr267@uky.edu*

Follow this and additional works at: [https://uknowledge.uky.edu/lin\\_facpub](https://uknowledge.uky.edu/lin_facpub)



Part of the [Linguistics Commons](#)

**Right click to open a feedback form in a new tab to let us know how this document benefits you.**

---

### Repository Citation

Tanner, James; Sonderegger, Morgan; Stuart-Smith, Jane; and Fruehwald, Josef, "Toward "English" Phonetics: Variability in the Pre-Consonantal Voicing Effect across English Dialects and Speakers" (2020). *Linguistics Faculty Publications*. 75.  
[https://uknowledge.uky.edu/lin\\_facpub/75](https://uknowledge.uky.edu/lin_facpub/75)

This Article is brought to you for free and open access by the Linguistics at UKnowledge. It has been accepted for inclusion in Linguistics Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

---

## Toward "English" Phonetics: Variability in the Pre-Consonantal Voicing Effect across English Dialects and Speakers

Digital Object Identifier (DOI)

<https://doi.org/10.3389/frai.2020.00038>

### Notes/Citation Information

Published in *Frontiers in Artificial Intelligence*, v. 3, article 38, p. 1-15.

© 2020 Tanner, Sonderegger, Stuart-Smith and Fruehwald.

This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Toward “English” Phonetics: Variability in the Pre-consonantal Voicing Effect Across English Dialects and Speakers

James Tanner<sup>1\*</sup>, Morgan Sonderegger<sup>1</sup>, Jane Stuart-Smith<sup>2</sup> and Josef Fruehwald<sup>3</sup>

<sup>1</sup> Department of Linguistics, McGill University, Montreal, QC, Canada, <sup>2</sup> Glasgow University Laboratory of Phonetics, University of Glasgow, Glasgow, United Kingdom, <sup>3</sup> Department of Linguistics, University of Kentucky, Kentucky, KY, United States

Recent advances in access to spoken-language corpora and development of speech processing tools have made possible the performance of “large-scale” phonetic and sociolinguistic research. This study illustrates the usefulness of such a large-scale approach—using data from multiple corpora across a range of English dialects, collected, and analyzed with the SPADE project—to examine how the pre-consonantal Voicing Effect (longer vowels before voiced than voiceless obstruents, in e.g., *bead* vs. *beat*) is realized in spontaneous speech, and varies across dialects and individual speakers. Compared with previous reports of controlled laboratory speech, the Voicing Effect was found to be substantially smaller in spontaneous speech, but still influenced by the expected range of phonetic factors. Dialects of English differed substantially from each other in the size of the Voicing Effect, whilst individual speakers varied little relative to their particular dialect. This study demonstrates the value of large-scale phonetic research as a means of developing our understanding of the structure of speech variability, and illustrates how large-scale studies, such as those carried out within SPADE, can be applied to other questions in phonetic and sociolinguistic research.

**Keywords:** voicing effect, English, phonetic variability, Bayesian modeling, dialectal variation, speaker variability

## OPEN ACCESS

### Edited by:

Jack Grieve,  
University of Birmingham,  
United Kingdom

### Reviewed by:

Joshua Waxman,  
Yeshiva University, United States  
George Bailey,  
University of York, United Kingdom

### \*Correspondence:

James Tanner  
james.tanner@mail.mcgill.ca

### Specialty section:

This article was submitted to  
Language and Computation,  
a section of the journal  
Frontiers in Artificial Intelligence

**Received:** 24 December 2019

**Accepted:** 01 May 2020

**Published:** 29 May 2020

### Citation:

Tanner J, Sonderegger M,  
Stuart-Smith J and Fruehwald J  
(2020) Toward “English” Phonetics:  
Variability in the Pre-consonantal  
Voicing Effect Across English Dialects  
and Speakers. *Front. Artif. Intell.* 3:38.  
doi: 10.3389/frai.2020.00038

## 1. INTRODUCTION

There exist a large number of well-studied properties of speech that are known to vary across languages and communities of speakers, which have long been of interest to sociolinguists and phoneticians. One dimension of this variability, which is the focus of this study, is that of variation *within languages*: across dialects and their speakers. For example, the deletion of word-final /t/ and /d/ segments (in e.g., *mist*, *missed*) has been shown to vary across a wide range of dialects and speech communities (e.g., Labov et al., 1968; Guy, 1980; Tagliamonte and Temple, 2005), as have the dialect-specific realization of English vowels (e.g., Thomas, 2001; Clopper et al., 2005; Labov et al., 2006), and variation in the degree of aspiration in English voiced and voiceless stops (e.g., Docherty, 1992; Stuart-Smith et al., 2015; Sonderegger et al., 2017). The study of this kind of variation provides a means of understanding the sources and structures of variability within languages: both in how particular dialects may systematically differ from each other, and how the variable realization of speech sounds maps to speakers’ cognitive representation of language and speech (Liberman et al., 1967; Lisker, 1985; Kleinschmidt, 2018). Despite decades of research, however, there is much we

do not know about the scope, extent, and structure of this kind of language-internal variability. Within the phonetic literature, most research has focused on highly-controlled speech styles in ‘laboratory settings,’ generally focusing on a single dialect in each study; much of the work focusing on phonetic variability in spontaneous speech is on single dialects (e.g., Ernestus et al., 2015). The sociolinguistic and dialectological literatures have often examined spontaneous speech, with some notable cross-dialectal studies (e.g., Clopper et al., 2005; Labov et al., 2006; Jacewicz and Fox, 2013), but nonetheless primarily focus on variation in vowel quality. Increasingly, however, research within phonetics and sociophonetics is being performed at a larger scale *across* speech communities (Labov et al., 2006, 2013; Yuan et al., 2006, 2007; Yuan and Liberman, 2014; Coleman et al., 2016; Liberman, 2018), driven by the development of new speech processing tools and data sharing agreements. This “large-scale” approach is applied here to one such well-studied variable, the pre-consonantal voicing effect, as a means of characterizing its degree and structure of variability in a single phonetic effect across English dialects and speakers.

The pre-consonantal voicing effect (henceforth *Voicing Effect*, VE) refers to vowels preceding voiced obstruents being consistently longer than their voiceless counterparts, such as the differences in *beat-bead* and *mace-maze* (House and Fairbanks, 1953; House, 1961). The VE has been reported—to greater or lesser extent—in a range of languages (Zimmerman and Sapon, 1958; Chen, 1970), though varies in size based on properties of the phonetic environment, such as whether the obstruent is a stop or fricative, the height of the vowel, and many others (Klatt, 1973; Crystal and House, 1982; Port and Dalby, 1982). The evidence for the English VE to date is sourced predominantly from laboratory studies of highly-controlled speech, often in citation form, recorded from small numbers of often standard General American English speakers (e.g., Rositzke, 1939; House and Fairbanks, 1953; Peterson and Lehiste, 1960; House, 1961; Crystal and House, 1982; Luce and Charles-Luce, 1985). On the basis of this evidence, the VE has been noted for being particularly large in English relative to other languages (Zimmerman and Sapon, 1958; Chen, 1970), and has long been suggested as a prominent cue to consonant voicing in English (Denes, 1955; Klatt, 1973). This in turn has motivated claims that the VE is learned in English, as opposed to being a low-level phonetic property in other languages (Fromkin, 1977; Keating, 2006; Solé, 2007). At the same time, numerous questions about the nature and extent of the VE in English remain unexplored. In this study, we will examine the variability in the VE across a range of English dialects, focusing on the following two research questions: (1) *how large is the VE as realized in spontaneous English speech?* and (2) *how much does the VE vary across dialects and speakers?* In addressing these questions, we hope to gain insight into a number of open issues, including the extent to which there is a single “English” VE or whether dialects differ in the magnitude of the effect, as well as the range of VE sizes across individual speakers of a given dialect.

This paper answers these questions by taking a “large-scale” approach to the study of the VE. Concretely, this refers to the use of a large amount of acoustic data, collected from a large number

of speakers across a range of English dialects. This analysis falls within the framework of the *SPeech Across Dialects of English* (SPADE) project (Sonderegger et al., 2019, <https://spade.glasgow.ac.uk/>), which aims to consider phonetic and phonological variation in British and North American English across time and space through the use of automated acoustic analysis of features across English dialects occurring in many corpora. The methodological and research goals of the SPADE project are exemplified through this study of the English VE, specifically by the use of multiple corpora of diverse sources and structures, and the use of linguistic and acoustic analysis via the *Integrated Speech Corpus Analysis* (ISCAN) tool (McAuliffe et al., 2019), developed as part of the broader SPADE project. Both the volume and complexity of the resulting data and the goals of the study motivate the need for appropriately-flexible approaches to the statistical analysis: specifically, the data is statistically analyzed using Bayesian regression models (Carpenter et al., 2017), which enable us to accurately estimate the size of the VE across dialects and speakers directly, whilst controlling for the complex nature of the spontaneous speech data.

The structure of this paper is as follows. Section 2 outlines previous work on the VE, and some of the outstanding questions related to our current understanding of its variability. Section 3 describes the data: the corpora of different dialects from SPADE. Sections 4, 5 describe the methodological approach: the process of acoustic and statistical analysis of the data. The results of this analysis are reported in section 6, and then discussed with respect to our specific research questions in section 7 and concluding in section 8.

## 2. THE VOICING EFFECT (VE)

The observation that vowels preceding voiced obstruents are consistently longer than before voiceless obstruents was first noted in early phonetics textbooks (e.g., Sweet, 1880; Kenyon, 1940; Thomas, 1947; Jones, 1948) and in preliminary experimental work from the first half of the twentieth century (Heffner, 1937; Rositzke, 1939; Hibbitt, 1948). Studies explicitly manipulating the VE in English observed an effect of around 1.45—that is, vowels before voiced consonants were longer than before voiceless consonants by a ratio of around 2:3 (House and Fairbanks, 1953; House, 1961), and this effect was a cue to the voicing of the obstruent (Denes, 1955; Lisker, 1957; Raphael, 1972).

In these studies, VE was shown to be affected by consonant manner: namely, that fricatives showed a smaller or minimal VE compared to stops (Peterson and Lehiste, 1960), and less-robustly cued the voicing of the final consonant (Raphael, 1972). Initial studies of connected speech suggested that the size of the VE in this type of speech is more variable: VEs in carrier sentences are similar to those in isolated words (Luce and Charles-Luce, 1985)<sup>1</sup> whilst vowels in read or spontaneous speech exhibit smaller VE sizes of around 1.2, and a negligible VE for fricatives (Crystal and House, 1982; Tauberer and Evanini, 2009). VE size

<sup>1</sup>Harris and Umeda (1974), in their study of overall vowel duration, attribute this difference to a “mechanical” prosody as a consequence of numerous repetitions.

is also modulated by the overall length of the vowel, which is hypothesized to be due to an intrinsic incompressibility of the vowel, limited by the minimal time required to perform the articulatory motor commands necessary for vowel production (Klatt, 1976). This general suggestion has been supported by observations that VE is smaller for unstressed and phrase-medial vowels (Umeda, 1975; Klatt, 1976), and vowels produced at a faster speech rate (Crystal and House, 1982; Cuartero, 2002). The VE is thus modulated by a range of phonetic factors, and largely predict a reduction of VE size in instances where vowels are generally shorter; vowels that undergo “temporal compression” have a reduced capacity to maintain a large VE size, and so VE is minimized. As these effects have only been investigated in laboratory speech, it is not clear whether the size and direction of these effects are maintained in less-controlled spontaneous speech styles.

Examining the VE across languages, Zimmerman and Sapon (1958) first observed that whilst English speakers produced a robust VE, Spanish speakers did not modulate vowel length in the same way, though this study did not control for the syllabic structure of test items. Comparing across English, French, Russian, and Korean, Chen (1970) observed that all four languages produced a VE size of at least 1.1, though all languages had different VE sizes (English = 1.63, French = 1.15, Russian = 1.22, Korean = 1.31). This was interpreted as evidence that VE is a phonetically-driven effect with additional language-specific phonological specification (Fromkin, 1977). Mack (1982), comparing English and French monolinguals with bilinguals, observed that English monolinguals maintained a substantially larger VE than French monolinguals, whilst the French-English bilinguals also produced the shorter French-style pattern instead of adapting to the larger English VE pattern. Keating (1985) suggested that VE is “phonetically-preferred,” though ultimately controlled by the grammar of the particular language. English, then, is expected to have a larger VE than other languages, though it is not known if the English VE is of a comparable size in spontaneous speech.

The work discussed above has not differentiated between varieties of English, and cross-linguistic comparisons of VE have presumed that a single “English” VE size exists. Little work has focused on variation in VE across English dialects beyond a small number of studies on specific dialects. One dialect group of interest has been Scottish Englishes and the application of the Scottish Vowel Length Rule (SVLR), where vowels preceding voiced fricatives and morpheme boundaries are lengthened, whilst all other contexts have short vowels (Aitken, 1981), and hence do not show the VE. In studies of the SVLR, some East Coast Scotland speakers show some evidence of the VE in production (Hewlett et al., 1999), whilst VE-like patterns were not observed in spontaneous Glaswegian (Rathcke and Stuart-Smith, 2016). On the other hand, studies of African American English (AAE) have claimed that voiced stops undergo categorical devoicing in this variety, which has resulted in additional vowel lengthening before voiced stops to maintain the pre-consonantal voicing contrast (Holt et al., 2016; Farrington, 2018). Only one study has previously compared the VE across English dialects in spontaneous speech. Tauberer and Evanini

(2009), using interview data from the *Atlas of North American English* (Labov et al., 2006), observe that North American English dialects vary in their VE values, ranging from 1.02 to 1.33, and that dialects with shorter vowels on average (New York City) also show a smaller-than-average VE size (1.13). Moreover, despite recognition that individual speakers may exhibit variability in their VE sizes (Rositzke, 1939; Summers, 1987), no study has formally examined the extent of variability across speakers, nor how dialects may differ in the degree of VE variability amongst its speakers. The two patterns observed for Scottish and African American English suggest that English dialects can maintain relatively “small” (or no), and “large” VEs, respectively; we know little about the degree of VE variability beyond these dialects without a controlled study across multiple English varieties, which is one of the goals of this study.

Whilst a large number of studies on the VE have provided useful information for its realization in English and other languages, there are still a range of outstanding questions that can be addressed through a large-scale cross-dialectal approach. To what extent is the VE a *learned* property of a given language, compared with an *automatic* consequence of low-level phonetic structure? Much of the discussion with respect to variation in VE has revolved around differences across *languages* (Chen, 1970; Keating, 1985), which may differ both in their phonetic realization of segments but also the phonological representation of those segments. In this sense, examining VE variability internal to a language (i.e., across *dialects*) potentially avoids this problem; the specification of phonological categories—here, the voicing status of final obstruents—are expected to be largely consistent within a language, meaning that language-internal variability may be driven by only differences in phonetic implementation.

Little is known about how English dialects may vary in their implementation of the VE, and so a range of possibilities exist for how dialects might compare. One possibility is that, with the exception of varieties with specific phonological rules interacting with the VE, dialects might cluster around a single “English” VE value, potentially of the size reported in the previous literature. Such a finding would support the previous approach in the literature, in terms of English compared to other languages, and suggest that dialects do not differ in how the final voicing contrast is phonetically implemented. Alternatively, dialects may differ gradually from each other, and so may show a continuum of possible dialect-specific VE sizes. If dialects do differ in their VE size in this way, this would suggest that the previous literature on the VE in “English” accounts for just a fraction of the possible VE realizations across English, and would provide evidence that individual English dialects differ in their phonetic implementation of an otherwise “phonological” contrast (Keating, 1984, 1985).

Similarly, little is known about how individual speakers vary in the VE, and what the overall distribution of speaker VE sizes is. Synchronic variability across speakers is one of the key inputs to sound change (Ohala, 1989; Baker et al., 2011), and also defines the limits of a speech community, i.e., speakers who share sociolinguistic norms in terms of production and social evaluation (e.g., Labov, 1972). Whilst dialects may differ in the realization of segments or the application of phonological

processes, dialect-internal variability is potentially more limited if a phonetic alternation such as the VE is critical to speech community membership.

### 3. DATA FOR THIS STUDY

The varieties of English included in this study are from North America, Great Britain, and Ireland. For the purposes of this study, North American dialects refer to the regions of the United States and Canada outlined in *The Atlas of North American English*, which is based around phonetic, not lexical, differences between geographic regions (Labov et al., 2006; Boberg, 2018). For Canadian data specifically, the primary distinction was made between “urban” and “rural” speakers, based on its relative importance noted in comparison to much weaker geographic distinctions, at least for the corpus which makes up most Canadian data in this study (Rosen and Skriver, 2015). Within the British and Irish groups, dialects from England in this study are defined in terms of Trudgill’s dialectal groupings (Trudgill, 1999), which groups regions in terms of both phonological and lexical similarity. Due to the lack of geographical metadata for speakers from Ireland and Wales, these dialects were simply coded as “Ireland” and “Wales” directly. Scottish Englishes are grouped based on information from *The Scottish National Dictionary*<sup>2</sup>. The data used in this study comes from the SPADE project, which aims to bring together and analyze over 40 speech corpora covering English speech across North America, the United Kingdom, and Ireland. In this study, we analyze data from 15 of these corpora, which together cover 30 different English dialects from these regions, comprised of speech from interviews, conversations, and reading passages. A basic description of each of these corpora is given below, outlining the type of speech and phonetic alignment tools used.

- *Audio British National Corpus* (AudioBNC, Coleman et al., 2012): The spoken sections of the British National Corpus, originally containing speech from over 1,000 speakers. However, due to a range of recording issues (e.g., overlapping speech, background noise, microphone interference), a large portion of the corpus is inaccurately aligned. In order to define a subset of the AudioBNC which maximizes the accuracy of the alignment, utterances were kept if they met a number of criteria: the utterance length was greater than one second, that the utterance contained at least two words, that the mean harmonics-to-noise ratio of the recording was at least 5.6, and that the mean difference in segmental boundaries between the alignment and a re-alignment with the Montreal Forced Aligner (MFA, McAuliffe et al., 2017a) was at most 30 ms<sup>3</sup>. 50 TextGrids from the remaining data were manually checked and deemed to be as approximately accurate as that of normal forced-alignment.
- *Brains in Dialogue* (Solanki, 2017): recordings of 24 female Glaswegian speakers producing spontaneous speech in a laboratory setting. There are 12 recordings for each speaker, which were aligned with LaBB-CAT (Fromont and Hay, 2012).
- *Buckeye* (Pitt et al., 2007): spontaneous interview speech of 40 speakers from Columbus Ohio, recorded in 1990s–2000s. The Buckeye corpus is hand-corrected with phonetic transcription labels: these were converted back to phonological transcriptions in order to be comparable with data from the other corpora.
- *Corpus of Regional African American Language* (CORAAAL, Kendall and Farrington, 2018): spontaneous sociolinguistic interviews with 100 AAE speakers from Washington DC, Rochester NY, and Princeville NC, recorded between 1968 and 2016, and aligned with the MFA.
- *Doubletalk* (Geng et al., 2013): recordings of paired speakers carrying out a variety of tasks in order to elicit a range of styles/registers in a discourse/interactive situation. Ten speakers make up five pairs where one member is a speaker of Southern Standard British English and the other member is a speaker of Scottish English.
- *Hastings* (Holmes-Elliott, 2015): recordings of sociolinguistic interviews with 46 speakers from Hastings in the south east of England, male and female, aged from 8 to 90, aligned using FAVE (Rosenfelder et al., 2014).
- *International Corpus of English—Canada* (ICE-Canada, Greenbaum and Nelson, 1996): interview and broadcast speech of Canadian English, recorded in the 1990s across Canada, and aligned using the MFA. Speaker dialect was defined in terms of their city or town of origin. In this study, we coded a speaker as “urban” if their birthplace was a large Canadian city.
- *Canadian Prairies* (Rosen and Skriver, 2015): Spontaneous sociolinguistic interviews, recorded between 2010 and 2016, with speakers of varying ethnic backgrounds from the provinces of Alberta and Manitoba, conducted as part of the Language in the Prairies project, and was aligned using the MFA.
- *Modern RP* (Fabricius, 2000): reading passages by Cambridge University students recorded in 1990s and 2000s. The speakers were chosen for having upper middle-class backgrounds as defined by at least one parent having a professional occupation along with the speaker also having attended private schooling. The data used in this study come from a reading passage aligned with FAVE.
- *Philadelphia Neighborhood Corpus* (PNC, Labov and Rosenfelder, 2011): sociolinguistic interviews with 419 speakers from Philadelphia, recorded between 1973 and 2013, and were aligned with FAVE.
- *Raleigh* (Dodsworth and Kohn, 2012): semi-structured sociolinguistic interviews of 59 White English speakers in Raleigh, North Carolina, born between 1955 and 1989, and aligned with the MFA.
- *Santa Barbara* (Bois et al., 2000): spontaneous US English speech, recorded in the 1990s and 2000s, from a range of speakers of different regions, genders, ages, and social backgrounds.

<sup>2</sup>Part of *The Dictionary of the Scots Language* (<https://dsl.ac.uk/>).

<sup>3</sup>We are grateful to Michael Goodale for designing and performing this filtering protocol.

- *The Scottish Corpus of Texts and Speech* (SCOTS, Anderson et al., 2007): approximately 1,300 written and spoken texts (23% spoken), ranging from informal conversations, interviews, etc. Most spoken texts were recorded since 2000.
- *Sounds of the City* (SOTC, Stuart-Smith et al., 2017): vernacular and standard Glaswegian from 142 speakers over 4 decades (1970s–2000s), collected from historical archives and sociolinguistic surveys, aligned using LaBB-CAT.
- *Switchboard* (Godfrey et al., 1992): 2,400 spontaneous telephone conversations between random participants from the multiple dialect regions in the United States on a variety of topics, containing data from around 500 speakers.

The goals of this study are to examine the size and variability in the English VE in spontaneous speech, and in variation in the VE across dialects and individual speakers. Specifically, the kind of dialectal variability being addressed in this study is that of *regional* variability: variability by race or ethnicity is not being directly considered in this study, with the exception of three African American English varieties, given the particular observations about AAE with respect to the VE (Holt et al., 2016; Farrington, 2018). This study also does not focus on differences according to age, either age-grading or apparent/real-time change in the VE over time; only speech data recorded since 1990s was included; the other data recorded prior to 1990 was excluded from further analysis. Analysis of the role of age and time in the VE in these English dialects remains a subject for future study.

## 4. DATA ANALYSIS

Having collected and organized the speech data into dialects, it is then possible to extract and acoustically analyze the data in the study: that is, going from raw data (audio and transcription files) to datasets which can be statistically analyzed. As the corpora differ in their formats—the phone labels used, organization of speaker data, etc.—modifying the acoustic analysis procedure for each different corpus format would be both labor and time-intensive, as well as increase the risk that the analysis itself differed across corpora. In order to standardize the acoustic analysis across corpora, the *Integrated Speech Corpus Analysis* (ISCAN) tool was developed for use in this kind of cross-dialectal study in the context of the SPADE project. This section provides a brief overview of the ISCAN system: see McAuliffe et al. (2017b, 2019) and the ISCAN documentation page for details of the implementation<sup>4</sup>.

The process of deriving a dataset from raw corpus files consists of three major steps. In the first step, individual speech corpora (in the form of sets of audio-transcription pairs) are *imported* into a graph database format, where each transcription file is minimally composed of word and phone boundaries (e.g., word-level and phone-level tiers in a TextGrid), and these word-phone relationships are structurally-defined in the database (i.e., that each phone belongs to a word). Importers have been developed for a range of standard automatic aligners,

including all formats of corpora described in section 3. Corpora, represented in database format, can then be further *enriched* with additional structure, measurements, and linguistic information. For example, utterances can be defined as groups of words (separated silence of a specified length, e.g., 150 ms), syllables can be defined as a property between groups of adjacent phones. Once the database has been enriched with utterance and syllable information, speech rate (often defined as syllables per second within an utterance) can be calculated and included in the database. Similarly, information about words (such as frequency) or speakers (such as gender, age, dialect etc.) can be added to the corpus from metadata files. Once a corpus has been sufficiently enriched with linguistic and acoustic information, it is then possible to perform a *query* on the corpus at a given level of analysis. This level of analysis refers to the level of the hierarchy on which the resulting datafile should use as the main level of observation, for example individual phones, syllables, or utterances. Filters can be applied to a query to restrict it to the particular contexts of interest, for example, including only syllables occurring at the right edge of an utterance, or vowels followed by a specific subset of phone types (e.g., obstruents). Finally, the resulting query can then be *exported* into a data format (currently CSV only) for further analysis.

Each corpus was processed using the ISCAN software pipeline, and then combined into a single “master” dataset, containing all phonetic, dialect, and speaker information from all of the analyzed corpora necessary to carry out the analysis of the VE below. As the vowel duration annotations from the corpora (except for Buckeye) were created via forced alignment with a minimum duration of 10 ms and a time-step of 30 ms, any token with a vowel duration below 50 ms was excluded from further study, as is common in acoustic studies of vowel formants to exclude heavily reduced vowels (Dodsworth, 2013; Fruehwald, 2013). To reduce the additional prosodic and stress effects on vowel duration, the study only included vowels from monosyllabic words occurring phrase-finally, where a phrase is defined as a chunk of speech separated by 150 ms of silence. Raw speech rate was calculated as syllables per second within a phrase, from which two separate speech rates were derived. First, a mean speech rate for each speaker was calculated, which reflects whether a speaker is a “fast” or “slow” speaker overall. From that mean speech rate, a local speech rate was calculated as the raw rate for the utterance subtracted from the given speaker’s mean. This local speech rate can be interpreted as how fast or slow that speaker produced the vowel within that particular phrase *relative* to their average speech rate (Sonderregger et al., 2017; Cohen Priva and Gleason, 2018). Word frequency was defined using the SUBTLEX-US dataset (Brybaert and New, 2009). The final dataset contained 229,406 vowel tokens (1,485 word types) from 1,964 speakers from 30 English dialects. **Table 1** shows the number of speakers and tokens for each dialect, and how many speakers/tokens were derived from each speech corpus.

## 5. STATISTICAL ANALYSIS

The research goals of this study focus on the size and variability of the VE in English spontaneous speech, and how the VE

<sup>4</sup><https://iscan.readthedocs.io/>

**TABLE 1** | Number of speakers and tokens per dialect (left), and by corpora from which each dialect was derived.

Region	Dialect	<i>n</i> Speakers	<i>n</i> tokens	Corpus	<i>n</i> speakers	<i>n</i> tokens
North America	Canada (rural)	52	9,313	Canadian Prairies	44	8,316
				ICE-Canada	8	997
	Canada (urban)	64	12,124	Canadian Prairies	56	11,939
				ICE-Canada	8	185
	Midwest US	40	5,567	Buckeye	40	5,567
	New England	24	1,336	Santa Barbara	7	174
				Switchboard	17	1,162
	North Midland US	46	3,084	Switchboard	46	3,084
	Northern Cities US	21	1,377	Santa Barbara	21	1,377
	Northern US	58	3,086	Switchboard	58	3,086
	NYC	25	1,477	Santa Barbara	6	158
				Switchboard	19	1,319
	Philadelphia	371	59,581	PNC	371	59,581
	Princeville NC (AAE)	71	6,759	CORAAL	17	6,759
	Raleigh US	92	3,282	Raleigh	92	3,282
	Rochester NY (AAE)	14	6,308	CORAAL	14	6,308
	South Midland US	108	8,188	Switchboard	108	8,188
	Southern US	44	2,738	Santa Barbara	6	345
				Switchboard	38	2,393
	Washington DC (AAE)	50	21,205	CORAAL	50	21,205
Western US	100	5,456	Santa Barbara	50	2,900	
			Switchboard	50	2,556	
United Kingdom & Ireland	Central Scotland	24	2,426	SCOTS	24	2,426
	East Central England	51	2,544	Audio BNC	51	2,544
				Doubletalk	5	726
	East England	229	20,727	Audio BNC	132	6,622
				Hastings	44	12,642
				ModernRP	48	737
				SCOTS	27	2,294
	Edinburgh	18	1,148	SCOTS	18	1,148
	Glasgow	177	33,938	Brains in Dialogue	23	9,210
				SCOTS	127	2,2434
				SOTC	127	2,2434
	Insular Scotland	8	351	SCOTS	8	351
	Ireland	19	624	Audio BNC	19	624
	Lower North England	60	3,325	Audio BNC	60	3,325
	North East England	17	488	Audio BNC	17	488
	Northern Scotland & Islands	33	2,280	SCOTS	33	2,280
	Scotland	70	3,468	Audio BNC	65	2,633
				Doubletalk	5	835
	South West England	50	2,067	Audio BNC	50	2,067
	Wales	41	2,524	Audio BNC	41	2,524
West Central England	41	2,615	Audio BNC	41	2,615	
Total		1,964	229,406			

varies across dialects and speakers. These goals motivate an approach of *estimating* the size of the VE in these contexts, rather than testing whether the VE “exists” or not. Whilst controlled laboratory experiments are explicitly designed to balance across these contexts (by including matching numbers of tokens with stops vs. fricatives, using words with similar

frequency, etc.), spontaneous speech taken from corpora is rarely balanced in this sense: some speakers speak more than others, have different conversations leading to some combinations of segments occurring infrequently relative to others, speakers manage properties of their speech (such as speech rate) for communicative purposes which are generally

absent in laboratory studies. In trying to obtain an accurate estimate of the VE (or indeed any other linguistic property), the unbalanced nature of spontaneous speech motivates the need for a statistical approach where individual factors of interest (e.g., obstruent manner of articulation, dialects, etc.) can be explored whilst controlling for the influence of other effects. This approach—the use of multiple regression to model corpus data—is now common in phonetics and sociolinguistic research (e.g., Tagliamonte and Baayen, 2012; Roettger et al., 2019), but has not, to our knowledge, been used to analyze multiple levels of variability in the VE.

In this study, this approach to estimation is performed using Bayesian regression modeling. Whilst other multifactorial statistical models would also be valid, Bayesian models provide us with some advantages that make the goal of estimating the size of the VE easier. Mixed-models are ideal for use in this study, as these capture variability at multiple levels (the VE overall, across dialects, across speakers) and this variability is of direct interest for our research questions. Bayesian mixed models resemble more traditional linear mixed-effects (LME) models approaches commonly used in linguistic and phonetic research, such as those performed with the *lme4* package (Bates et al., 2015), though differ in a few key respects. First, Bayesian models make it easy to calculate the *range* of possible VE sizes in each context, as opposed to a single value that would be output in LME models: whilst LME models provide ranges for “fixed” effects (across all dialects/speakers), Bayesian models provide a range of possible sizes for each level (i.e., an individual dialect). In a Bayesian model, all parameters (coefficients) in the model are assumed to have a *prior* distribution of possible values, reflecting which effect sizes are believed to be more or less likely, before examining the data itself. The output of a Bayesian model is a set of *posterior* distributions, which result from combining the priors and the likelihood of observing the data. Each model parameter has its own posterior distribution, which each represent the range of values for that parameter that is consistent with both the modeled data, conditioned on prior expectations about likely values, and the structure of the model itself. Bayesian models are well-suited to the task in this study, as they allow for flexible fitting of model parameters, and allow the complex random-effects structures which are often recommended for fitting statistically-conservative models (Barr et al., 2013), but which often fail to converge in LME models (Nicenboim and Vasishth, 2016). See Vasishth et al. (2018) for an introduction to Bayesian modeling applied to phonetic research.

A Bayesian mixed model of log-transformed vowel duration was fit using *brms* (Bürkner, 2018): a R-based front-end for the Stan programming language (Carpenter et al., 2017), containing the following population-level (“fixed effects”) predictors: the **voicing** and **manner** of the following obstruent, vowel **height** (high vs. non-high), the lexical **class** of the word (lexical vs. functional), both **mean** and **local** speech rates, and lexical **frequency**. To observe how compression of the vowel influences VE size, interactions between all of these factors with obstruent voicing were also included. The continuous predictors (both speech rates, frequency), were centered and divided by two standard deviations (Gelman and Hill, 2007). The two-level

factors (obstruent voicing, manner, vowel height, lexical class) were converted into binary (0,1) values and then centered.

The group-level (“random effects”) structure of the model contained the complete set of model predictors for both dialects and speakers, nested within dialects. These terms capture two kinds of variability in the VE size: for each individual dialect, as well as the degree of variability across speakers—the nesting of speaker term inside dialects can be interpreted as capturing the variability in the size of the VE across speakers *within* a given dialect. Given the expectation that both the overall vowel duration (represented by the intercept) and the manner of the obstruent would affect the size of the VE, correlation terms between the intercept and both the consonant voicing and manner predictors, as well as for the interaction *between* the voicing and manner predictors, were included for both dialects and speakers. Random intercepts were included for words and phoneme labels, also nested within dialects. The model was fit using 8,000 samples across 4 Markov chains (2000/2000 warmup/sample split per chain) and was fit with weakly informative “regularizing” priors (Nicenboim and Vasishth, 2016; Vasishth et al., 2018): the intercept prior used a normal distribution with a mean of 0 and a standard deviation of 1 [written as *Normal*(0, 1)]; the other fixed effects parameters used *Normal*(0, 0.5) priors, with the exception of the obstruent voicing parameter which used a *Normal*(0.1, 0.2) prior<sup>5</sup>. The group-level (for dialects, speakers) parameters used the *brms* default prior of a half Student’s *t*-distribution with 3 degrees of freedom and a scale parameter of 10. The correlations between group-level effects used the LKJ (Lewandowski et al., 2009) with  $\zeta = 2$ , which gives lower prior probability to perfect (−1/1) correlations, as recommended by Vasishth et al. (2018).

## 6. RESULTS

The results in this study will be reported in the context of the two main research questions concerning VE variability (1) in spontaneous speech, and (2) across English dialects and individual speakers. The results are reported for each effect in terms of the median value with 95% credible intervals (CrIs), and the probability of that effect’s direction. These values enable us to understand the *size* of the effect (i.e., the change in vowel duration) and the confidence in the effect’s predicted direction. The strength of evidence for an effect is distinct from the strength of the effect itself: to value the strength of evidence for an effect, we follow the recommendations of Nicenboim and Vasishth (2016) and consider there to be *strong* evidence of an effect if the 95% credible interval does not include 0, and *weak* evidence for an effect if 0 is within the 95% CrI but the probability of the effect’s direction is at least 95% (i.e., that there is <5% probability that the effect changes direction). Evaluating the strength of an effect

<sup>5</sup>The values chosen for the obstruent voicing parameter reflect the decision to allow a wide range of possible VE sizes, including values both above and below those reported in the previous literature. A sensitivity analysis was performed using an additional model fit with a “uniform” flat prior for the obstruent voicing parameter, which returned VE values differing by an order of  $10^{-3}$ , suggesting that the decision for the weakly-informative prior did not adversely affect the reported results.

**TABLE 2** | Posterior mean ( $\hat{\beta}$ ), estimated error, upper & lower credible intervals, and posterior probability of the direction of each population-level parameter included in the model of log-transformed vowel duration.

Parameter	$\hat{\beta}$	Est.Error	95% CrI	Pr( $\hat{\beta} <> 0$ )
Intercept	-1.99	0.02	[-2.03, -1.96]	1
Obstruent voicing	0.14	0.03	[0.09, 0.19]	1
Obstruent manner	0.05	0.02	[0.02, 0.08]	1
Vowel height	-0.22	0.02	[-0.25, -0.18]	1
Lexical class	-0.14	0.03	[-0.21, -0.08]	1
Speech rate (mean)	-0.22	0.01	[-0.24, -0.20]	1
Speech rate (local)	-0.28	0.01	[-0.30, -0.26]	1
Lexical frequency	-0.05	0.01	[-0.08, -0.03]	1
Voicing : Manner	-0.04	0.03	[-0.10, 0.02]	0.91
Voicing : Height	0.07	0.02	[0.02, 0.11]	1
Voicing : Class	-0.07	0.03	[-0.13, 0.00]	0.97
Voicing : Mean rate	-0.01	0.01	[-0.03, 0.01]	0.77
Voicing : Local rate	-0.06	0.01	[-0.08, -0.03]	1
Voicing : Frequency	-0.07	0.02	[-0.11, -0.03]	1

is determined with respect to effect sizes previously reported for laboratory (e.g., House and Fairbanks, 1953; House, 1961) and connected speech (Crystal and House, 1982; Tauberer and Evanini, 2009). The degree of variability across dialects can be compared with the findings of Tauberer and Evanini (2009); as there is no known comparison for speaker variability, this will be compared to variability across dialects as an initial benchmark.

## 6.1. The Voicing Effect in Spontaneous Speech

Table 2 reports the population-level (“fixed”) effects for each parameter in the fitted model. The “overall” VE size averaging across dialects, which is between 1.09 and 1.2, is estimated to be smaller than reported in previous laboratory studies ( $\hat{\beta} = 0.14$ , CrI = [0.09, 0.19],  $\text{Pr}(\hat{\beta} > 0) = 1$ )<sup>6</sup> and more consistent with VE sizes reported in studies of connected and spontaneous speech (Crystal and House, 1982; Tauberer and Evanini, 2009).

Looking at how the overall VE size for all dialects is modulated by phonetic context, there is weak evidence that the manner of the following obstruent modulates VE size ( $\hat{\beta} = -0.04$ , CrI = [-0.10, 0.02],  $\text{Pr}(\hat{\beta} < 0) = 0.91$ ): whilst stops appear to have a larger VE size (Figure 1, top left), the uncertainty in VE size for each obstruent manner (represented by the spread of the credible intervals) suggests that it is possible there is no difference in VE size between both obstruent manners. Whilst high vowels are shown to be shorter than non-high vowels overall ( $\hat{\beta} = -0.22$ , CrI = [-0.25, -0.18],  $\text{Pr}(\hat{\beta} < 0) = 1$ ), there is strong evidence that high vowels have a larger VE than non-high vowels ( $\hat{\beta} = 0.07$ , CrI = [0.02, 0.11],  $\text{Pr}(\hat{\beta} > 0) = 1$ ). There is a similarly strong effect for lexical class ( $\hat{\beta} = -0.07$ , CrI = [-0.13, 0.00],  $\text{Pr}(\hat{\beta} < 0) = 0.97$ ), where functional words have smaller VEs than

open-class lexical items (Figure 1, top right). Lexical frequency also has a strong and evident effect on VE size ( $\hat{\beta} = -0.07$ , CrI = [-0.11, -0.03],  $\text{Pr}(\hat{\beta} < 0) = 1$ ), where higher-frequency words have smaller VEs than their lower-frequency counterparts (Figure 1, bottom left), whilst local speech rate also reduces VE size ( $\hat{\beta} = -0.06$ , CrI = [-0.08, -0.03],  $\text{Pr}(\hat{\beta} < 0) = 1$ ; Figure 1, bottom middle). For mean speaking rate, however, the effect on VE is both small with weak evidence ( $\hat{\beta} = -0.01$ , CrI = [-0.03, 0.01],  $\text{Pr}(\hat{\beta} < 0) = 0.77$ ): this is reflected in Figure 1 (bottom right), where the difference between faster and slower speakers has a negligible effect on VE size. These results generally suggest that shorter vowels (within-speaker) tend to have smaller VE sizes, consistent with the temporal compression account (Klatt, 1973): the apparent exception to this is the relationship between VE size and vowel height, which is addressed in section 7.

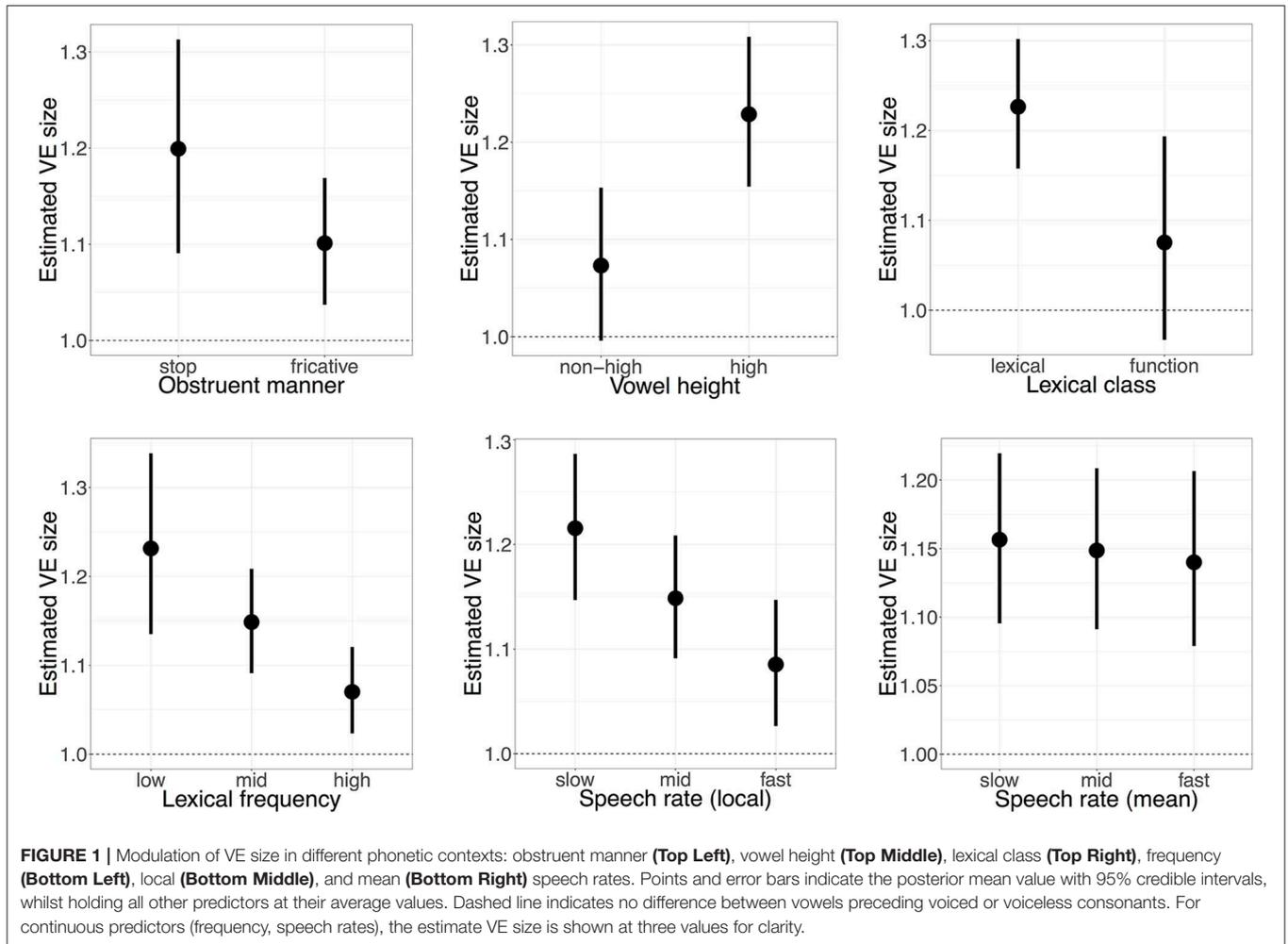
## 6.2. Voicing Effect Across Dialects and Speakers

Turning to dialectal variability in VE, we observe that the dialect variation in VE (the dialect-level standard deviation,  $\hat{\sigma}_{\text{dialect}}$ ) is between 0.07 and 0.12: this can be interpreted as meaning that the difference in VE size between a “low” and “high” VE dialect is between 32 and 61%<sup>7</sup> (Table 3). This is comparable with the range of possible values for the overall VE (between 0.09 and 0.19, Table 2, row 2). To understand whether this constitutes a “large” degree of variability, one metric is to assess whether a “low VE” dialect would actually have a reversed effect direction (voiceless > voiced), which is tested by subtracting  $2 \times \hat{\sigma}_{\text{dialect}}$  from the overall VE size and comparing to 0. There is little evidence that dialects differ enough to change direction ( $\hat{\beta} = -0.05$ , CrI = [-0.09, 0],  $\text{Pr}(\hat{\beta} > 0) = 0.06$ ), which suggests that whilst individual dialects differ in the size of the VE, no dialect fully differs in the direction of the effect (i.e., no dialect’s credible interval is fully negative).

Another way of understanding the degree of dialectal variability in VE is to examine the predicted VE for individual dialects. As shown in Figure 2, dialects appear to differ gradually from each other, ranging from dialects with effectively-null VE to those with strong evidence for large VEs. The Scottish dialects of Central Scotland and Edinburgh have VEs of at most 1.06 and 1.09, respectively, based on their upper credible interval value, whilst their median values (indicated by the points in Figure 2) indicate that the most likely VE size is around 0 (Central Scotland:  $\hat{\beta} = 0.99$ , CrI = [0.93, 1.06]; Edinburgh:  $\hat{\beta} = 1.01$ , CrI = [0.93, 1.09]): indeed, all Scottish dialects have a predicted VE size of 1.16 at the highest, with most of these having median values <1.1 (Table 4). North American dialects, in contrast, all have robustly positive VE values (no credible interval crosses the 0 line) and are generally larger than the British and Irish variants, shown by the position of red (North American) and blue (United Kingdom and Ireland) points respectively in Figure 2. In particular, the AAE dialects have the largest VEs in the sample, which are all robustly larger than the average “English” VE size (Rochester NY:  $\hat{\beta} = 1.35$ , CrI = [1.27, 1.44]; Princeville NC:  $\hat{\beta} =$

<sup>6</sup>As vowel duration was log-transformed prior to fitting, effects are interpreted by taking the exponent of the model parameter’s value, e.g.,  $e^{0.19} = 1.2$ , which refers to a vowel duration increase of 20%.

<sup>7</sup>The value is multiplied by 4 to get the 95% range of values =  $2\hat{\sigma}_{\text{dialect}}$  for both sides of the distribution = 0.28, which is then back-transformed from log via the exponential function =  $e^{0.28} = 1.32$ .



**FIGURE 1 |** Modulation of VE size in different phonetic contexts: obstruent manner (**Top Left**), vowel height (**Top Middle**), lexical class (**Top Right**), frequency (**Bottom Left**), local (**Bottom Middle**), and mean (**Bottom Right**) speech rates. Points and error bars indicate the posterior mean value with 95% credible intervals, whilst holding all other predictors at their average values. Dashed line indicates no difference between vowels preceding voiced or voiceless consonants. For continuous predictors (frequency, speech rates), the estimate VE size is shown at three values for clarity.

**TABLE 3 |** Posterior mean ( $\hat{\sigma}$ ), estimated error, and 95% credible intervals for dialect and speaker-level parameters related to obstruent voicing included in the model of log-transformed vowel duration.

Level	Parameter	$\hat{\sigma}$	Est.Error	95% CrI
Dialect	Intercept	0.05	0.01	[0.03, 0.07]
	Obstruent Voicing	0.09	0.01	[0.07, 0.12]
	Voicing : Manner	0.12	0.02	[0.09, 0.16]
	Voicing : Height	0.04	0.01	[0.01, 0.06]
	Voicing : Class	0.06	0.01	[0.04, 0.09]
	Voicing : Mean Rate	0.02	0.01	[0.00, 0.05]
	Voicing : Local Rate	0.05	0.01	[0.03, 0.07]
Speaker	Intercept	0.10	0.00	[0.09, 0.10]
	Obstruent Voicing	0.08	0.00	[0.07, 0.08]
	Voicing : Height	0.11	0.01	[0.10, 0.12]
	Voicing : Manner	0.11	0.01	[0.10, 0.13]
	Voicing : Class	0.13	0.01	[0.11, 0.14]
	Voicing : Local Rate	0.09	0.01	[0.08, 0.11]

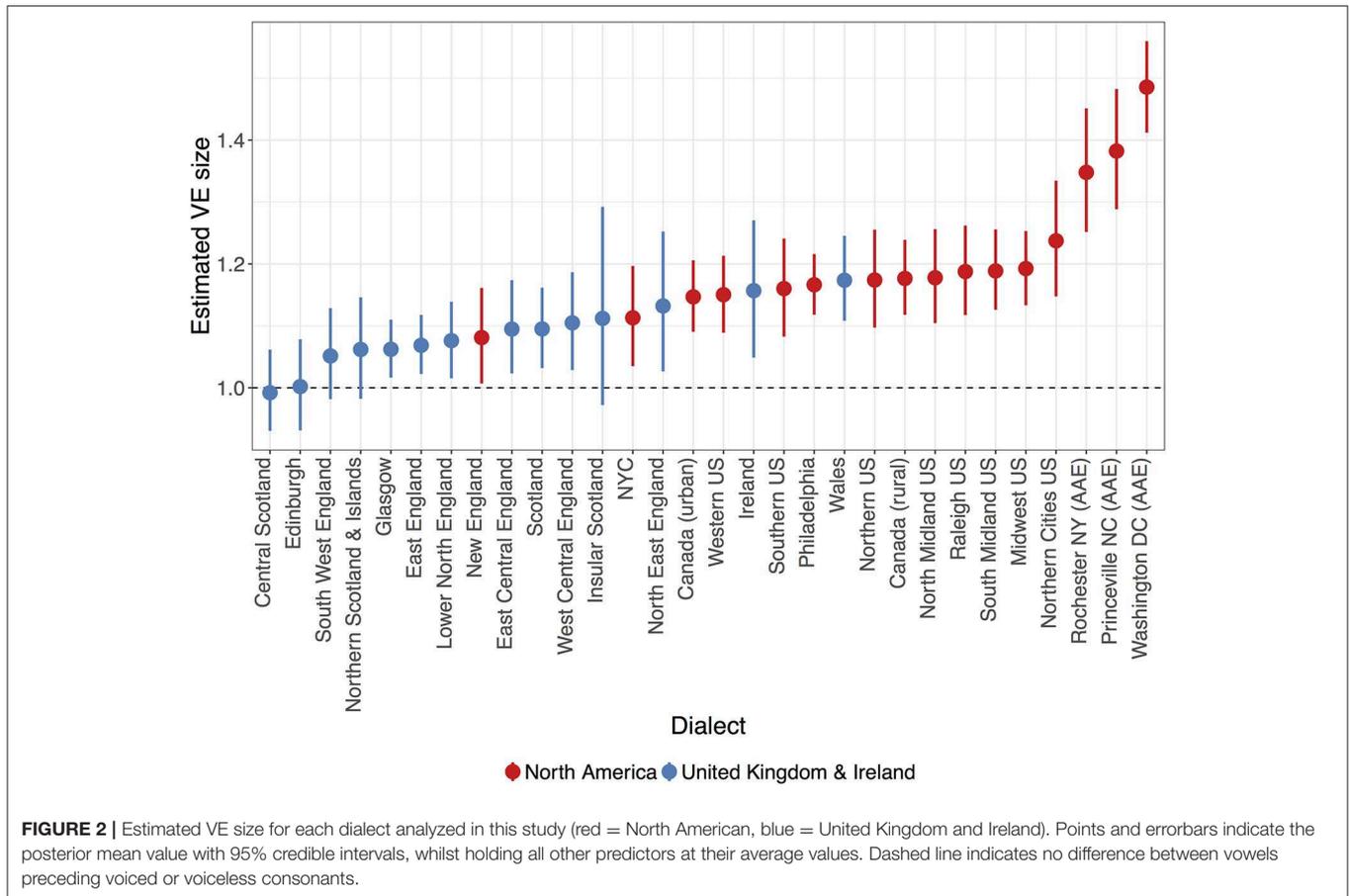
1.39, CrI = [1.31, 1.48]; Washington DC:  $\hat{\beta} = 1.49$ , CrI = [1.42, 1.56]): this is consistent with previous studies of studies on AAE,

which posit that final devoicing of word-final voiced obstruents results in compensatory vowel lengthening (Holt et al., 2016; Farrington, 2018).

Turning to variability in VE across individual speakers, we observe that speakers are estimated to vary within-dialect by between 0.07 and 0.08 ( $\hat{\sigma}_{speaker} = 0.08$ , CrI = [0.07, 0.08]), meaning that speakers differ in their VE ratios by between 32 and 37% (Table 3). To put this value in context and get an impression of the size of variability across speakers, this value is compared with the degree of variability across dialects. Figure 3 illustrates how likely the model deems different degrees of by-speaker and by-dialect variability: highest probability (darker shading) lies where by-dialect variability is greater than by-speaker variability. By the metric of between-dialect variability, Figure 3 illustrates that whilst dialects differ in VE size, individual speakers vary little from their dialect-specific baseline value.

## 7. DISCUSSION

The findings from this study will be discussed with respect to the two research questions: (1) how the VE is realized in spontaneous speech, and (2) how the VE varies across dialects and speakers. The VE in English is often considered to be substantially larger



than in other languages (Chen, 1970) and claimed to play a significant perceptual role in cueing consonant voicing (Denes, 1955). Taken together, these observations have formed the basis for claims that the VE in English is phonologically specified beyond an otherwise phonetically-consistent acoustic property across languages (Fromkin, 1977; Keating, 1985). Previous work has focused on controlled laboratory speech, leaving open the question of how the VE is realized in spontaneous English speech.

In this study, the overall VE in spontaneous speech was observed to have a maximum size of around 1.2—substantially smaller than the 1.5 commonly reported in laboratory studies (e.g., House and Fairbanks, 1953; Peterson and Lehiste, 1960; House, 1961; Chen, 1970), and more consistent with previous research on VE in connected speech (Crystal and House, 1982; Tauberer and Evanini, 2009). Spontaneous VE size was also shown to be affected by a range of phonetic factors, such as consonant manner, vowel height, frequency, and speech rate, though the evidence for each of these effects varies substantially (section 6.1). What the effects of these phonetic factors suggest is that contexts where vowels are often shorter also have shorter VE sizes, supporting the argument of “temporal compression”: that vowels which have already undergone shortening cannot be subsequently shortened further (Harris and Umeda, 1974; Klatt, 1976). An interesting exception to this finding is that the VE size was found to be larger for high vowels than non-high

vowels in this study (**Figure 1**)—the direction of this effect may be counter to that predicted by temporal compression, and opens a question as to whether this and other predictions of temporal compression are straightforwardly replicable in spontaneous speech environments. The overall smaller-size and impact of phonetic factors of the VE in spontaneous speech indicates a possible fragility of the VE in spontaneous speech, in apparent contrast to the supposed perceptual importance of the VE as a cue to consonant voicing (Denes, 1955; Lisker, 1957; Raphael, 1972). This apparent conflict between the perceptual importance of the VE and its subtlety in production provides an interesting area for future work.

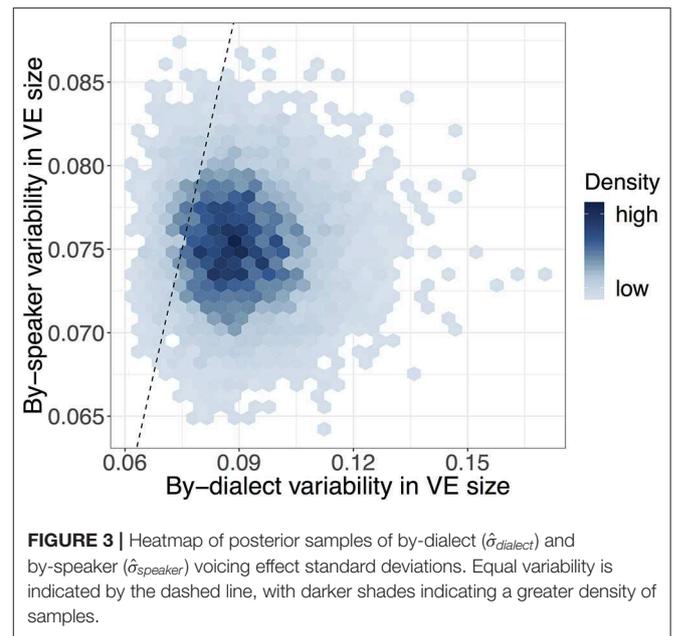
The fact that VE size in English differs so widely between laboratory and connected speech not only demonstrates the importance of speech style and context on phonetic realization (Labov, 1972; Lindblom, 1990), but also raises the question of “how big” the VE in English really is, or could be. If larger overall VE size is only observable in laboratory speech, it would be interesting to empirically re-evaluate the question of whether English VE is in fact larger than in other languages. For languages that exhibit smaller VEs than English in laboratory speech (Chen, 1970), it is not clear how such languages may realize the VE in more naturalistic speech. One possibility is that the VE across languages is comparatively small in spontaneous speech and similarly affected by phonetic factors; alternatively, the VE in

**TABLE 4** | Estimated VE sizes (mean, estimated error, and upper and lower credible intervals) for each dialect used in this study.

Dialect	$\hat{\beta}$	Est.Error	95% CrI
Central Scotland	0.99	0.03	[0.93, 1.06]
Edinburgh	1.01	0.04	[0.93, 1.09]
South West England	1.05	0.03	[0.99, 1.12]
Glasgow	1.06	0.02	[1.02, 1.11]
Northern Scotland & Islands	1.06	0.04	[0.99, 1.14]
East England	1.07	0.02	[1.02, 1.12]
Insular Scotland	1.08	0.06	[0.96, 1.21]
Lower North England	1.08	0.03	[1.02, 1.15]
New England	1.08	0.04	[1.00, 1.17]
East Central England	1.09	0.03	[1.03, 1.16]
Scotland	1.10	0.03	[1.04, 1.16]
West Central England	1.11	0.03	[1.04, 1.18]
NYC	1.12	0.04	[1.04, 1.20]
North East England	1.14	0.05	[1.04, 1.26]
Canada (urban)	1.15	0.02	[1.09, 1.21]
Western US	1.15	0.03	[1.09, 1.21]
Canada (rural)	1.17	0.03	[1.12, 1.24]
Ireland	1.17	0.04	[1.07, 1.28]
Philadelphia	1.17	0.02	[1.12, 1.22]
Southern US	1.17	0.03	[1.10, 1.24]
North Midland US	1.18	0.03	[1.11, 1.26]
Northern US	1.18	0.03	[1.11, 1.26]
Wales	1.18	0.03	[1.11, 1.25]
Raleigh US	1.19	0.03	[1.13, 1.26]
South Midland US	1.19	0.03	[1.13, 1.26]
Midwest US	1.20	0.03	[1.14, 1.26]
Northern Cities US	1.24	0.04	[1.15, 1.33]
Rochester NY (AAE)	1.35	0.03	[1.27, 1.44]
Princeville NC (AAE)	1.39	0.03	[1.31, 1.48]
Washington DC (AAE)	1.49	0.02	[1.42, 1.56]

spontaneous speech across other languages may still be smaller than in English and retain cross-linguistic differences akin to those reported by Chen (1970), and thus English would still retain its status as a language with a distinct realization of the VE.

The first research question (section 6.1) considered how the VE was modulated in spontaneous speech, averaging across dialects. To what extent dialects themselves differ in VE was the focus of the second research question. As shown in section 6.2, English was shown to exhibit a range of different VE sizes across individual dialects. The dialects with the smallest and largest VEs—Scottish Englishes and AAE, respectively—were expected to show these values given evidence of additional phonological rules governing vowel duration in these varieties (Aitken, 1981; Holt et al., 2016; Rathcke and Stuart-Smith, 2016; Farrington, 2018). Beyond these varieties, dialects appear to differ gradually from each other, ranging in VE values from around 1.05 in South West England to 1.24 in the Northern Cities region (Figure 2). As opposed there being a single “English” VE value, there appears to be a range of VE sizes within the language. Such a finding further complicates the notion that English has a particular and large



VE relative to other languages. Imagining these different dialects as “languages” with minimally different phonological structures, this finding demonstrates that such similar “languages” can have very different phonetic effects (Keating, 1985). This in turn underlies a more nuanced approach to the question of whether English truly differs from other languages in its VE size: not only may English have varieties with greater or lesser VE sizes, but other languages may also exhibit similar dialectal VE ranges.

Individual speakers are also shown to vary in the realization of the VE, though the extent of this variability is rather limited when compared to variability across dialects (Figure 3): that is, whilst dialects appear to demonstrate a range of possible VE patterns, individual speakers vary little from their dialect-specific baseline values. Such a finding supports an interpretation where the VE has a dialect-specific value which speakers learn as part of becoming a speaker of that speech community. The limited extent of speaker variability could predict that the VE will be stable within individual English dialects, given the key role of synchronic speaker variability as the basis for sound change (Ohala, 1989; Baker et al., 2011). This would need checking on a dialect-by-dialect basis, however, given recent evidence of Glaswegian undergoing weakening in its vowel duration patterns (Rathcke and Stuart-Smith, 2016). It also highlights the need for studies addressing both synchronic and diachronic variability across dialects, which we hope to address in future work. One important caveat to the finding is that it assumes that all the dialects analyzed in this study contain only speakers who are speakers of that dialect: if a given dialect had a particularly large degree of by-speaker variability, it could be that this could reflect the existence of multiple speakers of different dialects (and thus different VE patterns) within that particular dialect coding. This is unlikely to be a particular problem in this study, however, as a separate model that allows for by-speaker variability to vary on a per-dialect basis showed that no dialect with a

sufficiently large number of tokens exhibited overly large by-speaker variability (section 6.2).

By using speech data from multiple sources and multiple dialects, it has been possible to investigate variability of a phonological feature across “English” overall, examine variability at the level of individual dialects and speakers, and reveal the extent of English-wide phonetic variability that was not previously apparent in studies of individual dialects and communities. In this sense, our “large-scale” approach, using consistent measures and controlling factors, enables us to understand the nature of dialectal variability in the English VE directly within the context of both other dialects and English as a whole.

Whilst this kind of study extends the scope of analysis for (socio)phonetic research, there are of course a number of limitations that should be kept in mind in studies of this kind. This study of the English VE predominantly uses data from automatic acoustic measurements, in turn calculated from forced aligned-segmented datasets. All forced-alignment tools have a minimum time resolution (often 10 ms), a minimum segment duration (often 30 ms), and there always exists the possibility of poor or inaccurate alignment. This is a necessary consequence of the volume of data used in this study: there is simply *too much* data to manually check and correct all durations, and so the best means of limiting these effects is through sensible filtering and modeling of the data. For example, segments with aligned durations of less than 50 ms were excluded, since accurately capturing the duration of a vowel this small could be difficult given the time resolution of the aligner. This decision could exaggerate the size of the VE estimation, as only the most reduced vowels have been removed from the data. Another property of forced alignment which impacts our study of VE is that aligners will only apply the phonological segment label to the segment, meaning that it is possible to only examine VE in terms of *phonological* voicing specification (i.e., whether a segment is underlyingly voiced or not), as opposed to whether the segment itself was realized with phonetic voicing. For example, the realization of the stop as devoiced (Farrington, 2018) or as a glottal stop (Smith and Holmes-Elliott, 2018), or the relative duration of the closure preceding the vowel (Lehiste, 1970; Port and Dalby, 1982; Coretta, 2019), could affect VE size which is not controllable by exclusively using phonological segment labels. How this kind of phonetic variation, and the more general relationship between a “phonological” and a “phonetic” VE, should be understood would certainly be an interesting project for future work. Finally, given the diversity of formats and structures of the corpora available for this study, it has only been possible to categorize and study dialects in a rather broad “regional” fashion. Similarly, we were unable to investigate the effect of speaker age due to the heterogenous coding of age across the corpora: we agree this is an important dimension that we have attempted to account for in the approach to statistical modeling, and is certainly necessary to examine in future work. Whilst these limitations may be less suitable for approaching other questions in phonetics and sociolinguistics which are concerned with variability at a more detailed level, the approach taken in this study points to a promising first step toward exposing the

structures underlying fine-grained phonetic variability at a larger level across multiple speakers and dialects of a language.

## 8. CONCLUSION

The recent increase in availability of spoken-language corpora, and development of speech and data processing tools have now made it easier to perform phonetic research at a “large-scale”—incorporating data from multiple different corpora, dialects, and speakers. This study applies this large-scale approach to investigate how the English Voicing Effect (VE) is realized in spontaneous speech, and the extent of its variability across individual dialects and speakers. Little has been known about how the VE varies across dialects bar a handful of studies of specific dialects (Aitken, 1981; Tauberer and Evanini, 2009; Holt et al., 2016). English provides an interesting opportunity to directly examine how phonetic implementation may differ across language varieties with minimally different phonological structures (Keating, 1985). By applying tools for automatic acoustic analysis (McAuliffe et al., 2019) and statistical modeling (Carpenter et al., 2017), it was found that the English VE is substantially smaller in spontaneous speech, as compared with controlled laboratory speech, and is modulated by a range of phonetic factors. English dialects demonstrate a wide degree of variability in VE size beyond that expected from specific dialect patterns such as the SVLR, whilst individual speakers are relatively uniform with respect to their dialect-specific baseline values. In this way, this study provides an example of how large-scale studies can provide new insights into the structure of phonetic variability of English and language more generally.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## AUTHOR CONTRIBUTIONS

JT extracted the data, performed the statistical analysis, and wrote the first draft of the manuscript. All authors contributed to the conception and design of the study. All authors contributed to manuscript revision, and read and approved the submitted version.

## FUNDING

The research reported here is part of SPeech Across Dialects of English (SPADE): Large-scale digital analysis of a spoken language across space and time (2017–2020); ESRC Grant ES/R003963/1, NSERC/CRSNG Grant RGPDD 501771-16, SSHRC/CRSH Grant 869-2016-0006, NSF Grant SMA-1730479 (Digging into Data/Trans-Atlantic Platform), and was also supported by SSHRC #435-2017-0925 awarded to MS and a Fonds de Recherche du Québec Société et Culture International Internship award granted to JT.

## ACKNOWLEDGMENTS

We acknowledge the crucial contribution of The SPADE Consortium (<https://spade.glasgow.ac.uk/the-spade-consortium/>), which comprises the Data Guardians who generously shared their datasets which are reported in this paper, and/or whose datasets were used in the development of the ISCAN tool. No SPADE research would have been possible without The

SPADE Consortium. This research reported in this paper is an extended version of a preliminary report in Tanner et al. (2019). We would like to thank the audiences of the 2019 Montreal-Ottawa-Toronto Phonology Workshop and UK Language Variation & Change 12 for their feedback on previous versions of this research. We also thank SPADE team members, especially Michael Goodale, Rachel Macdonald, and Michael McAuliffe for assistance with data management.

## REFERENCES

- Aitken, A. J. (1981). *The Scottish Vowel Length Rule. The Middle English Dialect Project*. Edinburgh.
- Anderson, J., Beavan, D., and Kay, C. (2007). “The Scottish corpus of texts and speech,” in *Creating and Digitizing Language Corpora*, eds J. C. Beal, K. P. Corrigan, and H. L. Moisl (New York, NY: Palgrave), 17–34. doi: 10.1057/9780230223936\_2
- Baker, A., Archangeli, D., and Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Lang. Variat. Change* 23, 347–374. doi: 10.1017/S0954394511000135
- Barr, D. J., Levy, R., Sheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Boberg, C. (2018). “Dialects of North American English,” in *Handbook of Dialectology*, eds C. Boberg, J. Nerbonne, and D. Watt (Oxford: John Wiley and Sons), 450–461. doi: 10.1002/9781118827628.ch26
- Bois, J. W. D., Chafe, W. L., Meyer, S. A., Thompson, S. A., and Martey, N. (2000). *Santa Barbara Corpus of Spoken American English*. Technical report, Linguistic Data Consortium, Philadelphia, PA.
- Brysbaert, M., and New, B. (2009). Moving beyond Kucera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav. Res. Methods* 41, 977–990. doi: 10.3758/BRM.41.4.977
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R J.* 10, 395–411. doi: 10.32614/RJ-2018-017
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* 76, 1–32. doi: 10.18637/jss.v076.i01
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22, 129–159. doi: 10.1159/000259312
- Clopper, C. G., Pisoni, D. B., and de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *J. Acoust. Soc. Am.* 118, 1661–1676. doi: 10.1121/1.2000774
- Cohen Priva, U., and Gleason, E. (2018). “The role of fast speech in sound change,” in *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (Austin, TX: Cognitive Science Society), 1512–1517.
- Coleman, J., Baghai-Ravary, L., Pybus, J., and Grau, S. (2012). *Audio BNC: The Audio Edition of the Spoken British National Corpus*. Technical report, Oxford. Available online at: <http://www.phon.ox.ac.uk/AudioBNC>
- Coleman, J., Renwick, M. E. L., and Temple, R. A. M. (2016). Probabilistic underspecification in nasal place assimilation. *Phonology* 33, 425–458. doi: 10.1017/S0952675716000208
- Coretta, S. (2019). An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish. *Glossa* 4, 1–25. doi: 10.5334/gjgl.869
- Crystal, T. H., and House, A. S. (1982). Segmental durations in connected speech signals: preliminary results. *J. Acoust. Soc. Am.* 72, 705–716. doi: 10.1121/1.388251
- Cuartero, N. (2002). *Voicing assimilation in Catalan and English* (Ph.D. thesis). Universitat Autònoma de Barcelona, Barcelona, Spain.
- Denes, P. (1955). Effect of duration on the perception of voicing. *J. Acoust. Soc. Am.* 27, 761–764. doi: 10.1121/1.1908020
- Docherty, G. (1992). *The Timing of Voicing in British English Obstruents*. Berlin; New York, NY: Foris. doi: 10.1515/9783110872637
- Dodsworth, R. (2013). Retreat from the Southern Vowel Shift in Raleigh, NC: social factors. *Univ. Pennsylvania Work. Pap. Linguist.* 19, 31–40. Available online at: <https://repository.upenn.edu/pwpl/vol19/iss2/5/>
- Dodsworth, R., and Kohn, M. (2012). Urban rejection of the vernacular: the SVS undone. *Lang. Variat. Change* 24, 221–245. doi: 10.1017/S0954394512000105
- Ernestus, M., Hanique, I., and Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *J. Phonet.* 38, 60–75. doi: 10.1016/j.wocn.2014.08.001
- Fabricius, A. H. (2000). *T-glottalling between stigma and prestige: a sociolinguistic study of Modern RP* (Ph.D. thesis). Copenhagen Business School, Copenhagen, Denmark.
- Farrington, C. (2018). Incomplete neutralization in African American English: the cast of final consonant devoicing. *Lang. Variat. Change* 30, 361–383. doi: 10.1017/S0954394518000145
- Fromkin, V. A. (1977). “Some questions regarding universal phonetics and phonetic representations,” in *Linguistic Studies Offered to Joseph Greenberg on the Occasion of His Sixtieth Birthday*, ed A. Juillard (Saratoga, NY: Anma Libri), 365–380.
- Fromont, R., and Hay, J. (2012). “LaBB-CAT: an annotation store,” in *Australasian Language Technology Workshop 2012, Vol. 113*, 113–117.
- Fruehwald, J. (2013). *The phonological influence on phonetic change* (Ph.D. thesis). University of Pennsylvania, Pennsylvania, PA, United States.
- Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511790942
- Geng, C., Turk, A., Scobbie, J. M., Macmartin, C., Hoole, P., Richmond, K., et al. (2013). Recording speech articulation in dialogue: evaluating a synchronized double electromagnetic articulography setup. *J. Phonet.* 41, 421–431. doi: 10.1016/j.wocn.2013.07.002
- Godfrey, J. J., Holliman, E. C., and McDaniel, J. (1992). “SWITCHBOARD: telephone speech corpus for research and development,” in *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing - Vol. 1* (San Francisco, CA), 517–520. doi: 10.1109/ICASSP.1992.225858
- Greenbaum, S., and Nelson, G. (1996). The International Corpus of English (ICE project). *World English*. 15, 3–15. doi: 10.1111/j.1467-971X.1996.tb00088.x
- Guy, G. (1980). “Variation in the group and the individual: the case of final stop deletion,” in *Locating Language in Time and Space*, ed W. Labov (New York, NY: Academic Press), 1–36.
- Harris, M., and Umeda, N. (1974). Effect of speaking mode on temporal factors in speech: vowel duration. *J. Acoust. Soc. Am.* 56, 1016–1018. doi: 10.1121/1.1903366
- Heffner, R.-M. S. (1937). Notes on the lengths of vowels. *Am. Speech* 12, 128–134. doi: 10.2307/452621
- Hewlett, N., Matthews, B., and Scobbie, J. M. (1999). “Vowel duration in Scottish English speaking children,” in *Proceedings of 14th The International Congress of Phonetic Sciences* (San Francisco, CA).
- Hibbitt, G. W. (1948). *Diphthongs in American speech: a study of the duration of diphthongs in the contextual speech of two hundred and ten male undergraduates* (Ph.D. thesis). Columbia University, New York, NY, United States.
- Holmes-Elliott, S. (2015). *London calling: assessing the spread of metropolitan features in the southeast* (Ph.D. thesis). University of Glasgow, Glasgow, Scotland.

- Holt, Y. F., Jacewicz, E., and Fox, R. A. (2016). Temporal variation in African American English: the distinctive use of vowel duration. *J. Phonet. Audiol.* 2. doi: 10.4172/2471-9455.1000121
- House, A. S. (1961). On vowel duration in English. *J. Acoust. Soc. Am.* 33, 1174–1178. doi: 10.1121/1.1908941
- House, A. S., and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. Acoust. Soc. Am.* 25, 105–113. doi: 10.1121/1.1906982
- Jacewicz, E., and Fox, R. A. (2013). “Cross-dialectal differences in dynamic formant patterns in American English vowels,” in *Vowel Inherent Spectral Change*, eds G. S. Morrison and P. F. Assmann (Berlin: Springer), 177–198. doi: 10.1007/978-3-642-14209-3\_8
- Jones, D. (1948). *An Outline of English Phonetics*. New York, NY: E. P. Dutton & Company.
- Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* 60, 189–218. doi: 10.2307/413642
- Keating, P. (2006). “Phonetic encoding of prosodic structure,” in *Speech Production: Models, Phonetic Processes, and Techniques*, eds J. Harrington and M. Tabain (New York, NY: Psychology Press), 197–186.
- Keating, P. A. (1985). “Universal phonetics and the organization of grammars,” in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, ed V. A. Fromkin (New York, NY: Academic Press), 115–132.
- Kendall, T., and Farrington, C. (2018). *The Corpus of Regional African American Language. Version 2018.10.06*, Eugene, OR.
- Kenyon, J. S. (1940). *American Pronunciation*. Ann Arbor, MI: George Wahr.
- Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *J. Acoust. Soc. Am.* 54, 1102–1104. doi: 10.1121/1.1914322
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *J. Acoust. Soc. Am.* 59, 1208–1221. doi: 10.1121/1.380986
- Kleinschmidt, D. F. (2018). Structure in talker variability: how much is there and how much can it help? *Lang. Cogn. Neurosci.* 34, 1–26. doi: 10.1080/23273798.2018.1500698
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Labov, W., Ash, S., and Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin: Mouton de Gruyter. doi: 10.1515/9783110167467
- Labov, W., Cohen, P., Robins, C., and Lewis, J. (1968). *A Study of the Non-Standard English of Negro and Puerto Rican Speakers in New York City*. Technical Report 1 & 2, Linguistics Laboratory, University of Pennsylvania.
- Labov, W., and Rosenfelder, I. (2011). “New tools and methods for very large scale measurements of very large corpora,” in *New Tools and Methods for Very-Large-Scale Phonetics Research Workshop*, Pennsylvania, PA.
- Labov, W., Rosenfelder, I., and Fruehwalf, J. (2013). One hundred years of sound change in Philadelphia: linear incrementation, reversal, and reanalysis. *Language* 89, 30–65. doi: 10.1353/lan.2013.0015
- Lehiste, I. (1970). Temporal organization of higher-level linguistic units. *J. Acoust. Soc. Am.* 48:111. doi: 10.1121/1.1974906
- Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *J. Multivar. Anal.* 100, 1989–2001. doi: 10.1016/j.jmva.2009.04.008
- Lieberman, M. (2018). Corpus phonetics. *Annu. Rev. Linguist.* 5, 91–107. doi: 10.1146/annurev-linguistics-011516-033830
- Lieberman, M. A., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279
- Lindblom, B. (1990). “Explaining phonetic variation: a sketch of the h&h theory,” in *Speech Production and Speech Modelling*, Vol. 4 of NATO ASI Series, eds W. J. Hardcastle and A. Marchal (Dordrecht: Kluwer Academic Publishers), 403–439. doi: 10.1007/978-94-009-2037-8\_16
- Lisker, L. (1957). Linguistic segments, acoustic segments and synthetic speech. *Language* 33, 370–374. doi: 10.2307/411159
- Lisker, L. (1985). The pursuit of invariance in speech signals. *J. Acoust. Soc. Am.* 77, 1199–1202. doi: 10.1121/1.392185
- Luce, P. A., and Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *J. Acoust. Soc. Am.* 78, 1949–1957. doi: 10.1121/1.392651
- Mack, M. (1982). Voicing-dependent vowel duration in English and French: monolingual and bilingual production. *J. Acoust. Soc. Am.* 71, 173–178. doi: 10.1121/1.387344
- McAuliffe, M., Coles, A., Goodale, M., Mihuc, S., Wagner, M., Stuart-Smith, J., et al. (2019). “ISCAN: A system for integrated phonetic analyses across speech corpora,” in *Proceedings of the 19th International Congress of Phonetic Sciences* (Melbourne, VIC).
- McAuliffe, M., Scolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017a). *Montreal Forced Aligner [computer program]*. Available online at: <https://montrealcorpus.tools.github.io/Montreal-Forced-Aligner/>
- McAuliffe, M., Stengel-Eskin, E., Socolof, M., and Sonderegger, M. (2017b). “Polyglot and Speech Corpus Tools: a system for representing, integrating, and querying speech corpora,” in *Proceedings of Interspeech 2017* (Stockholm). doi: 10.21437/Interspeech.2017-1390
- Nicenboim, B., and Vasisht, S. (2016). Statistical methods for linguistic research: foundational ideas - part II. *Lang. Linguist. Compass* 10, 591–613. doi: 10.1111/lnc3.12207
- Ohala, J. (1989). “Sound change is drawn from a pool of synchronic variation,” in *Language Change: Contributions to the Study of Its Causes*, eds L. E. Breivik and E. H. Jahr (Berlin: Mouton de Gruyter), 173–198.
- Peterson, G. E., and Lehiste, I. (1960). Duration of syllable nuclei in English. *J. Acoust. Soc. Am.* 32, 693–703. doi: 10.1121/1.1908183
- Pitt, M. A., Dille, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). *Buckeye Corpus of Spontaneous Speech, 2nd Edn*. Columbus, OH: Ohio State University.
- Port, R. F., and Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Percept. Psychophys.* 32, 141–152. doi: 10.3758/BF03204273
- Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *J. Acoust. Soc. Am.* 51, 1296–1303. doi: 10.1121/1.1912974
- Rathcke, T., and Stuart-Smith, J. (2016). On the tail of the Scottish Vowel Length Rule in Glasgow. *Lang. Speech* 59, 404–430. doi: 10.1177/0023830915611428
- Roettger, T. B., Winter, B., and Baayen, R. H. (2019). Emergent data analysis in phonetic sciences: towards pluralism and reproducibility. *J. Phonet.* 73, 1–7. doi: 10.1016/j.wocn.2018.12.001
- Rosen, N., and Skriver, C. (2015). Vowel patterning of Mormons in Southern Alberta, Canada. *Lang. Commun.* 42, 104–115. doi: 10.1016/j.langcom.2014.12.007
- Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H., et al. (2014). *FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2* 10.5281/zenodo.22281.
- Rositzke, H. A. (1939). Vowel-length in General American speech. *Am. Speech* 15, 99–109. doi: 10.2307/408728
- Smith, J., and Holmes-Elliott, S. (2018). The unstoppable glottal: tracking rapid change in an iconic British variable. *English Lang. Linguist.* 22, 323–355. doi: 10.1017/S1360674316000459
- Solanki, V. J. (2017). *Brains in dialogue: investigating accommodation in live conversational speech for both speech and EEG data* (Ph.D. thesis). University of Glasgow, Glasgow, Scotland.
- Solé, M.-J. (2007). “Controlled and mechanical properties in speech,” in *Experimental Approaches to Phonology*, eds P. Beddor and M. Ohala (Oxford: Oxford University Press), 302–321.
- Sonderegger, M., Bane, M., and Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language* 93, 598–640. doi: 10.1353/lan.2017.0038
- Sonderegger, M., Stuart-Smith, J., McAuliffe, M., Macdonald, R., and Kendall, T. (2019). “Managing data for integrated speech corpus analysis in SPEECH Across Dialects of English (SPADE),” in *Open Handbook of Linguistic Data Management*, eds A. Berez-Kroeker, B. McDonnell, E. Koller, and L. Collister (Cambridge: MIT Press).
- Stuart-Smith, J., Jose, B., Rathcke, T., MacDonald, R., and Lawson, E. (2017). “Changing sounds in a changing city: an acoustic phonetic investigation of real-time change over a century of Glaswegian,” in *Language and a Sense of Place: Studies in Language and Region*, eds C. Montgomery and E. Moore (Cambridge: Cambridge University Press), 38–65. doi: 10.1017/9781316162477.004
- Stuart-Smith, J., Sonderegger, M., Rathcke, T., and Macdonald, R. (2015). The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Lab. Phonol.* 6, 505–549. doi: 10.1515/lp-2015-0015

- Summers, W. V. (1987). Effects of stress and final consonant voicing on vowel production: articulatory and acoustic analyses. *J. Acoust. Soc. Am.* 82, 847–863. doi: 10.1121/1.395284
- Sweet, H. (1880). *A Handbook of Phonetics*. London: MacMillan & Co.
- Tagliamonte, S., and Temple, R. (2005). New perspectives on an ol variable: (t, d) in British English. *Lang. Variat. Change* 17, 281–302. doi: 10.1017/S0954394505050118
- Tagliamonte, S. A., and Baayen, R. H. (2012). Models, forests, and trees of York English: was/were variation as a case study for statistical practice. *Lang. Variat. Change* 24, 135–178. doi: 10.1017/S0954394512000129
- Tanner, J., Sonderegger, M., Stuart-Smith, J., and SPADE-Consortium (2019). Vowel duration and the voicing effect across English dialects. *Univ. Toronto Work. Pap. Linguist.* 41, 1–13. doi: 10.33137/twpl.v41i1.32769
- Tauberer, J., and Evanini, K. (2009). “Intrinsic vowel duration and the post-vocalic voicing effect: some evidence from dialects of North American English,” in *Proceedings of Interspeech*.
- Thomas, C. K. (1947). *An Introduction to the Phonetics of American English*. New York, NY: Ronald Press Company.
- Thomas, E. R. (2001). *An Acoustic Analysis of Vowel Variation in New World English*. American Dialect Society.
- Trudgill, P. (1999). *The Dialects of England*. Oxford: Blackwell.
- Umeda, N. (1975). Vowel duration in American English. *J. Acoust. Soc. Am.* 58, 434–445. doi: 10.1121/1.380688
- Vasishth, S., Nicenboim, B., Beckman, M., Li, F., and Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: a tutorial introduction. *J. Phonet.* 71, 147–161. doi: 10.1016/j.wocn.2018.07.008
- Yuan, J., and Liberman, M. (2014). F0 declination in English and Mandarin broadcast news speech. *Speech Commun.* 65, 67–74. doi: 10.1016/j.specom.2014.06.001
- Yuan, J., Liberman, M., and Cieri, C. (2006). “Towards an integrated understanding of speaking rate in conversation,” in *Proceedings of Interspeech 2006*, Pittsburgh, PA.
- Yuan, J., Liberman, M., and Cieri, C. (2007). “Towards an integrated understanding of speech overlaps in conversation,” in *Proceedings of the International Congress of Phonetic Sciences XVI* (Saarbrücken), 1337–1340.
- Zimmerman, S. A., and Sapon, S. M. (1958). Note on vowel duration seen cross-linguistically. *J. Acoust. Soc. Am.* 30, 152–153. doi: 10.1121/1.1909521

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Tanner, Sonderegger, Stuart-Smith and Fruehwald. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.