Vowel duration and the voicing effect across English dialects

James Tanner[1], *Morgan Sonderegger*[1], *Jane Stuart-Smith*[2], and *The SPADE Data Consortium*[2]

[1]*Department of Linguistics, McGill University*

[2]*Glasgow University Laboratory of Phonetics, University of Glasgow*

June 10, 2019

The ‘voicing effect’—the durational difference in vowels preceding voiced and voiceless consonants—is a well-documented phenomenon in English, where it plays a key role in the production and perception of the English final voicing contrast. Despite this supposed importance, little is known as to how robust this effect is in spontaneous connected speech, which is itself subject to a range of linguistic factors. Similarly, little attention has focused on variability in the voicing effect across dialects of English, bar analysis of specific varieties. Our findings show that shown that the voicing of the following consonant exhibits a weaker-than-expected effect in spontaneous speech, interacting with manner, vowel height, speech rate, and word frequency. English dialects appear to demonstrate a continuum of potential voicing effect sizes, where varieties with dialect-specific phonological rules exhibit the most extreme values. The results suggest that the voicing effect in English is both substantially weaker than previously assumed in spontaneous connected speech, and subject to a wide range of dialectal variability.

1 Introduction

A well-established property across most dialects of English is the ‘voicing effect’ (VE), where vowels preceding voiceless consonants are shorter than those preceding their voiceless counterparts. The size of this effect in English, suggested to be of a ratio between 1.2 to 1.6 (House, 1961; House and Fairbanks, 1953; Tauberer and Evanini, 2009), is both substantially larger in English than other languages (Chen, 1970) and larger than otherwise explainable by purely articulatory properties (Ohala, 1983). In spite of these observations, much of this previous research has focused on VE in controlled laboratory speech. As a result, it is not clear how *robust* the VE is—that is, whether the VE operates as a constant effect in the same direction across phonological and phonetic contexts (e.g., consonant manner, vowel height, speech rate). One possibility is that the VE is robust across contexts, though the size of the effect has been shown to be at least partially modulated by the effects of speaking style (Port, 1977) and speech rate (Cuartero, 2002, in Solé, 2007). An alternative possibility is that positive VE is context-dependent: i.e., that a strong VE can be observed in some sets of environments and is absent in others. For example, it is possible that VE is neutralised in fast speech, where there may some upper limit on the extent to which a syllable can shortened, resulting in some form of 'maximal compression' of a vowel (Peterson and Lehiste, 1960; Klatt, 1973, 1976; Luce and Charles-Luce, 1985; Summers, 1987). Similarly, high frequency words may both contribute to shortening (leading to maximal compression) and reduce the need for the contrastive effect of VE. Thus, the focus of the first research question concerns understanding how each of these factors interact and modulate VE in spontaneous speech, and examining the robustness of the VE when these factors are accounted for.

A second question here concerns how robust the VE is *across* varieties of English. With the exception of studies into specific varieties, dialectal differences in VE have received relatively little attention in the previous literature. Indeed, the vast majority of previous research on English VE has focused on North American English, elicited in word lists or planned carrier sentences (e.g., House and Fairbanks, 1953; House, 1961; Umeda, 1975; Hewlett et al., 1999; Holt et al., 2016). The studies which have looked at dialectal variability have also looked exclusively at North American English (Jacewicz et al., 2007; Tauberer and Evanini, 2009). Both of these studies report small but significant dialect differences in the size of the VE, but with all dialects exhibiting positive VEs (i.e., ratios larger than 1). It is possible to conclude from this that, modulo varieties with dialect-specific phonological duration patterns related to consonant voicing, the size and direction of the VE across dialects is quantitatively similar, and would be consistent with a view that most dialects cluster around a single English-specific VE value. In what ways might dialects quantitatively differ from each other? Is it possible that dialects can be clustered into varieties that exhibit a robust VE and those that do not? This would be not unlike suggestions made for cross-linguistic differences in VE, where e.g., Polish is distinct from English in its lack of a consistent VE (Keating, 1985). Alternatively, dialects may simply differ gradiently from each other, where "English" exhibits a continuum of possible VE values. The second focus of this research, then, is to examine how VE is realised across English dialects.

2 Background

2.1 *Vowel duration*

Durational differences across vowels are partly constrained by the physical properties of the acoustic system: for example, high vowels are shorter crosslinguistically than their non-high counterparts (Solé and Ohala, 2010). At the same time, however, variation in vowel duration is clearly phonologically conditioned, as evidenced by significant differences in vowel durations across English dialects and languages (Keating, 1985; Wilson and Chodroff, 2017). Vowel duration is also heavily affected by properties of the prosodic environment, such as lexical and phrasal stress (Crystal and House, 1990; Umeda, 1975), distance from a phrasal boundary (Wightman et al., 1992), the number of segments in a word (Crystal and House, 1990), and speech rate (Crystal and House, 1990). The relative 'informativeness' of a speech unit—predictability, in terms of word frequency (Bybee, 2001), conditional probability (Jurafsky et al., 2001), etc—is correlated with the general reducability of segments, where more contextually informative words are less likely to undergo phonetic reduction than their less informative counterparts (Priva, 2017). More predictable units are inversely informative (less frequent words are *more* informative), which in turn results in more predictable words being more likely to be reduced, resulting in more centralised formants and shorter durations (Ayelett and Turk, 2006; Gahl, 2008; Ernestus and Warner, 2011).

2.2 *The voicing effect (VE)*

As noted above, the VE refers to the tendency of vowels preceding voiceless consonants to be shorter than those preceding their voiced counterparts, e.g., *beat* [bit] vs *bead* [biːd]. The VE has been intensively studied in American English laboratory speech since the 1950s, where it has been observed regardless of vowel or consonantal quality (House, 1961; House and Fairbanks, 1953), and that the relative duration of the vowel provides a strong perceptual cue to the voicing of the consonant (Denes, 1955; Raphael, 1972; Luce and Charles-Luce, 1985). Whilst the voicing effect has been observed in a range of languages, it is noticeably larger in English: Chen (1970), using data from single-word elicitation, reports that English vowels are approximately 1.63 times longer preceding voiced consonants, compared to 1.15 for French, 1.22 for Russian, and 1.3 for Korean. This suggests that a prospective phonetic effect may be subject to different implementation across different languages' phonological systems, where it has potentially undergone phonologisation within English (Klatt, 1976; Fromkin, 1977; Keating, 1985). Studies looking at the VE in connected speech—situating the word of interest within a larger linguistic context—have shown VE size to be generally smaller than in studies where the target word is produced in isolation (Harris and Umeda, 1974; Port, 1977). The VE has also been shown to be modulated by phonetic and prosodic factors: House and Fairbanks (1953) observed that both the voicing and manner of the following consonant affected duration (where following fricatives and nasals results in longer vowels than following stops), but did not examine the interaction between voicing and manner. Examining the relationship between the VE and syllable stress, and it was shown that unstressed syllables show a smaller VE effect than stressed syllables (Klatt, 1973). Cuartero (2002, in Solé, 2007) demonstrated that slower speaking rates resulted in a larger VE for English but not in Catalan, further underlining the different possible language-specific implementations of the VE, and suggesting the importance of speech rate for VE size.

Given the substantial variability in vowel realisation across dialects of English, it is reasonable to consider whether VE also exhibits dialectal variability. There are few empirical studies concerning variability in VE across English varieties, which have predominantly focused on dialectal variation within North American English. In the largest cross-dialectal study, Tauberer and Evanini (2009) investigate variability in the size of the VE in 12 dialects across North America using telephone interview data from *The Altas of North American English* (Labov et al., 2006). They observe an average VE ratio of 1.21, with VE values ranging from 1.02 to 1.33: crucially, even the largest value reported by Tauberer and Evanini is smaller than 1.63 reported by Chen (1970), further suggesting that VE in connected speech is smaller than in single-word elicitations. As the effects of consonant manner, intrinsic vowel duration, and speech rate are not controlled together in their analysis statistically, it is possible that some of the observed variability is related to both dialectal differences and the previously mentioned phonetic effects, potentially affecting what is or is not interpreted as a dialectal difference. Jacewicz et al. (2007), examining data from 54 speakers across Ohio, Wisconsin, and North Carolina, also report both a significant VE effect and that the size of the significantly differed across three US English dialects, though point out that this effect is small relative to other phonetic factors (such as vowel height). Holt et al. (2016), examining vowel variability in African American English (AAE) speakers in North Carolina, show that vowels preceding voiced stops are longer in AAE than White American English (WAE) speakers in order to account for variable devoicing of final voiced consonants: for AAE speakers, the cue for final voicing has shifted to vowel duration to compensate for the neutralisation of voicing of the stop (Farrington, 2018). As it is suggested that this duration compensation is triggered by the devoicing of stops, it is not clear whether equivalent voiced-voiceless durational differences should

Table 1: Summary information of corpora used in this study

Corpus	Dialect	*n* speakers (*n* female)	*n* tokens
Buckeye (Pitt et al., 2007)	Midwest	40 (20)	7933
CORAAL (Kendall and Farrington, 2018)	Washington DC (AAE)	50 (26)	22922
ICE-Can (Greenbaum and Nelson, 1996)	Canada	11 (1)	742
Modern RP (Fabricius, 2000)	RP	48 (24)	703
Raleigh (Dodsworth and Kohn, 2012)	Raleigh	98 (49)	3378
Santa Barbara (Bois et al., 2000)	Eastern New England	10 (4)	193
	Lower South	6 (1)	368
	Northern Cities	21 (7)	1440
	NYC	7 (3)	170
	West	55 (34)	3041
SCOTS (Anderson et al., 2007)	Central	24 (14)	2666
	Edinburgh	18 (8)	1236
	Insular	9 (7)	384
	Northern	28 (14)	1994
	Glasgow	27 (15)	2445
SOTC (Stuart-Smith et al., 2017)	Glasgow	46 (20)	8956
Total	15	498 (247)	58571

be expected for fricatives or affricates. Beyond differences in North American Englishes, Scottish varieties exhibit phonological processes that interact directly with the realisation of VE: the Scottish Vowel Length Rule (SVLR) involves the lengthening of vowels preceding voiced fricatives and morpheme boundaries, whilst all other contexts exhibit short vowels (Aitken, 1981). In studies of spontaneous Glaswegian speech, robust VE-style patterns have not been detected (Rathcke and Stuart-Smith, 2016), whilst East Coast Scottish exhibits durational patterns closer to the VE in single-word elicitations (Hewlett et al., 1999). Together, these studies suggest that there may in fact be substantial language-internal (i.e., cross-dialectal) variability in the size and structure of the VE. What is still unclear, however, is whether these dialects represent distinct phonetic implementations, or rather are on different ends of a continuum of possible VEs in English.

3 Methods

3.1 Data

This study forms a part of the larger *SPeech Across Dialects of English* (SPADE) project,[1] which aims to examine variability and stability across dialects of English, focusing on the structure of segmental varia-

[1] SPADE Project website: https://spade.glasgow.ac.uk/

tion synchronically and diachronically. This is achieved through the large-scale analysis of speech corpora comprising a range of British and North American Englishes, where variation can be investigated in terms of cross-dialectal patterns of both variability and stability across Englishes as a whole. Understanding the structure of the VE in English (the goal of this study) is one such example of the goals of the larger SPADE project. For this study, data from the following 8 corpora were used:

- *Buckeye* (Pitt et al., 2007): conversational speech of 40 speakers from Columbus Ohio, recorded in 1990s-2000s.
- *Corpus of Regional African American Language* (CORAAL) (Kendall and Farrington, 2018): sociolinguistic interviews with 100 AAE speakers from Washington DC, recorded between 1968 and 2016.
- *International Corpus of English - Canada* (ICECAN) (Greenbaum and Nelson, 1996): interview and broadcast speech of Canadian English, recorded in the 1990s across Canada.
- *Modern RP* (Fabricius, 2000): reading passages by Cambridge University students recorded in 1990s and 2000s.
- *Raleigh* (Dodsworth and Kohn, 2012): semi-structured sociolinguistic interviews of 59 WAE speakers in Raleigh, North Carolina. Speakers were either born before 1955 (group 1), between 1955 and 1978 (group 2), or between 1979 and 1989 (group 3).
- *Santa Barbara* (Bois et al., 2000): Naturally-occurring US English speech, recorded 1990s-00s, from a range of speakers of different regions, genders, ages, and social backgrounds.
- *The Scottish Corpus of Texts and Speech* (SCOTS) (Anderson et al., 2007): approximately 1300 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.

Data was extracted from each of these corpora using the *Integrated Speech Corpus ANalysis* (ISCAN) tool (McAuliffe et al., 2019), a open source software system for performing phonetic analysis across multiple corpora, in spite of their heterogenous annotations and formats. Each corpora was separately imported into ISCAN, wherein each annotated segment is represented as a node within a hierarchical graph structure: each phone is associated with its corresponding word (and thus its preceding and following segments), and results in a database of individual phones from that corpus. This database can then be enriched with additional linguistic (e.g., stress, frequency), acoustic (e.g., speech rate, VOT), and speaker-level (e.g., dialect, age, gender) information. This database can then be queried at a particular linguistic (e.g., phone) level, at which filters can be applied (e.g., utterance-final, monosyllabic), restricting the phones included in the query. The result of this query can then be exported as a separate CSV. Once CSVs had been derived from all of the corpora, the set of CSV files were merged into a single 'master' dataset, which was used for the statistical analysis. In this study, utterances were defined as units of speech separated by at least 150ms of silence.

Whilst most of the corpora used contain speech from a single dialect, SCOTS and Santa Barbara contain speech from several dialects. For these cases, it was necessary to cluster speakers from dialect regions, referring to existing dialect groupings, e.g. for Scottish English, we referred to the Scottish National Dictionary.[2] After grouping, regions with 5 or fewer speakers were excluded in order to provide more reliable estimates for dialects kept in the sample. As ethnicity is expected to play a large role in particular VE patterns within certain dialects (Holt et al., 2016; Farrington, 2018), having speakers of differing ethnicities within the same dialect group could result in misleading VE values. Thus, all non-white AE speakers were excluded from Santa Barbara (72 speakers), leaving the Washington DC speakers from the CORAAL corpus as the only AAE speakers in the analysis. In order to partially reduce the effect of speaker age on vowel duration, recordings made earlier than the 1980s were excluded, resulting in the removal of 91 speakers (12 from SOTC and 79 from CORAAL).

As the vowel duration measurements in each corpus are produced via forced alignment with a resolution of 10ms and a minimum duration of 30ms (with the exception of Buckeye, which was aligned manually), any tokens below 49ms were excluded to minimise noise from highly-reduced vowels, as is common in phonetic studies of vowel formants (e.g., Dodsworth, 2013; Fruehwald, 2013). In order to minimise the additional effects of lexical stress and differences in phrasal position, only monosyllabic words occurring utterance-finally were extracted from the corpora. Thus, it is possible to focus on a select set of phonetic and phonological factors (voicing, manner, height, speech rate). Raw speech rate was calculated as syllables per second within the utterance, from which two separate speech rates were calculated. First, a *mean* speech

[2] The Scottish National Dictionary is accessible as part of *The Dictionary of the Scots Language* (https://dsl.ac.uk/).

rate for each speaker, and a *local* speech rate: each token's deviation from that speaker's mean rate. These can be interpreted as representing the effect of being a fast versus slow speaker (mean), and how fast or slow a speaker is talking relative to their regular speaking rate (local). Word frequency was calculated using SUBTLEX-US (Brysbaert and New, 2009). In total, the data examined here contains 58571 tokens (1233 types) observed across 498 speakers (247 female) from 15 British and North American dialects. Summary information of each corpus can be seen in Table 1.

3.2 Model

A Bayesian mixed-effects linear regression of log-transformed vowel duration was fit in R (R Core Team, 2018) using *brms* (Bürkner, 2018), which provides an R interface for the Stan programming language (Stan Development Team, 2018): a log transformation was applied to satisfy the assumption of linear regression that the dependent variable can take on any real-numbered value. In this model, the fixed effects values reflect the effect of each predictor on English *overall*: for an "average speaker" from an "average dialect". As the questions pursued here are directly interested in variability across dialects, the particular effect on each dialect is represented within the random effects structure—specifically, how the effect of each parameter deviates from the overall fixed effect value. Within a Bayesian framework, uncertainties in random effects parameters are more straightforward to compute than in a frequentist *lmer*-style model, motivating the use of a Bayesian model. As dialectal and speaker variability are structured as random effects in the model reported here, examining the degree of uncertainty for each dialect's VE parameter is directly relevant to RQ2. For practical purposes, this model is similar to a *lmer*-style mixed-effects regression: the key differences are that parameters have prior and posterior probability distributions, and that the uncertainties in each random effect term (posterior distributions) can be examined. Another important aspect of a Bayesian model, for our purposes, will be that "credible intervals" can be easily computed for any parameter—the range of values in which we can have 95% confidence the parameter lies. In contrast, standard information returned by an *lmer* model (estimate, SE, "confidence intervals") can only be used to assess whether a parameter is different from 0—not what possible values it could take on.

Two-level factors (consonant voicing, manner, vowel height) were scaled and centred at 0 as numerical predictors, and continuous factors (mean and local speech rates, frequency) were centred and divided by two standard deviations. This standardisation makes the coefficients in the model comparable (Gelman and Hill, 2007), and can be interpreted as their effect on log-transformed vowel duration whilst all other predictors are held at their average levels. The fixed-effects structure for the model contained predictors for following consonant **voicing** (i.e., the VE), **manner**, vowel **height**, **mean** and **local** speech rates, and word **frequency** (log-transformed at word-level). In order to examine how each of these properties modulate VE (RQ1), the fixed-effects structure also contained two-way interaction terms for all predictors (manner, height, speech rate, frequency) with consonant voicing. As RQ2 is focused on dialectal variation in the VE, the random-effects structure for dialects and speakers mirrors that of the fixed-effects. Random effects for speakers were *nested* within dialects: each speaker is treated as belonging to a single dialect, and so speaker-level effects reflect the within-dialect variability among speakers of that dialect. In order to control for the established relationship between vowel duration and VE size (Crystal and House, 1982; Luce and Charles-Luce, 1985) and that VE in some dialects (e.g., Scottish Englishes, AAE) is conditioned by consonant manner (Holt et al., 2016; Rathcke and Stuart-Smith, 2016), correlations between the intercept, voicing, manner, and the voicing-manner interaction were included for both dialects and speakers. Random intercepts were also included for words and vowel phone labels, the latter of which was also nested within dialects to reflect the expectation that the realisation of individual vowels will substantially differ across dialects.

The default minimally-informative priors from *brms* were used for all model parameters except random effect correlations; these were given an LKJ prior (Lewandowski et al., 2009) with $\zeta = 2$, in order to give lower prior probability to perfect (1/-1) correlations, as recommended by Vasishth et al. (2018). The results are reported using the model parameters (i.e., change in log-transformed vowel duration) with 95% credible intervals (CrIs), and the posterior probability that the model parameter does not change direction, which acts as an informal p-value. A Bayesian model captures degrees of belief; of particular interest here is whether there is evidence for an effect at all (analogously to a "$p < 0.05$" decision rule). Following Paape et al. (2017), we consider there to be evidence for a non-zero effect if 0 is not included in the 95% CrI, and there to be weak evidence for such an effect if the 95% CrI includes 0 but the probability of the parameter being either positive or negative is above 95% (e.g., the model is 95% confident that the parameter is positive).

4 Results

4.1 RQ1: robustness of the VE across phonetic and phonological contexts

Table 2 reports the population-level effects of each parameter in the model. There is an effect of the following consonant on vowel duration (i.e., the VE), where vowels with voiced consonants were longer than their

Table 2: Means ($\hat{\beta}$), Error, 95% credible intervals, and the probability of the posterior distribution including 0 (Pr($\hat{\beta} <> 0$)).

Parameter	$\hat{\beta}$	Error	5% CrI	95% CrI	Pr($\hat{\beta} <> 0$)
Intercept	-1.98	0.03	-2.04	-1.92	1.00
Voicing	0.08	0.04	0.01	0.15	0.98
Manner	0.05	0.03	0.00	0.10	0.97
Height	-0.14	0.03	-0.20	-0.09	1.00
Speech rate (mean)	-0.20	0.02	-0.23	-0.17	1.00
Speech rate (local)	-0.28	0.03	-0.33	-0.23	1.00
Frequency	-0.05	0.02	-0.09	-0.01	1.00
Voicing : Manner	-0.10	0.07	-0.24	0.03	0.94
Voicing : Height	-0.07	0.03	-0.13	0.00	0.98
Voicing : SR (mean)	-0.03	0.03	-0.09	0.02	0.89
Voicing : SR (local)	-0.06	0.02	-0.10	-0.01	0.99
Voicing : Frequency	-0.07	0.03	-0.13	-0.02	1.00

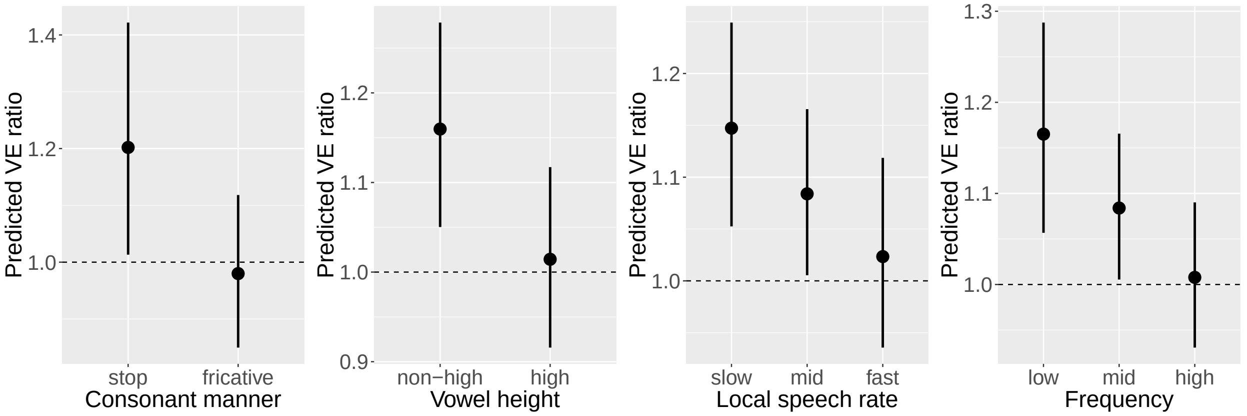

Figure 1: Predicted voicing effect (median and 95% CrI) as a function of consonant manner, vowel height, local speech rate (at -1, 0, 1), and word frequency (at -1, 0, 1). Predictions based on regression lines computed from the model's posterior, marginalising over all other covariates.

voiceless counterparts ($\hat{\beta}$ = 0.08, CrI = [0.01,0.15], Pr($\hat{\beta} > 0$) = 0.98). It should be noted, however, that the effect is predicted to be small: taking the exponent of the model coefficient and CrIs as the VE size, the model has 95% confidence that the effect is between 1 (e^{0}) and 1.16 ($e^{0.15}$) (median 1.08: $e^{0.08}$), compared with 1.63 for Chen (1970) in word list reading, and slightly smaller than the 1.21 for spontaneous North American English (Tauberer and Evanini, 2009). Whilst vowel duration is modulated by consonant manner ($\hat{\beta}$ = 0.05, CrI = [0.00,0.10], Pr($\hat{\beta} > 0$) = 0.97), Figure 1 (far left) illustrates that there is only weak evidence that the VE size is modulated by consonant manner, with a larger VE before stops than before fricatives ($\hat{\beta}$ = -0.10, CrI = [-0.24,0.03], Pr($\hat{\beta} < 0$) = 0.94). In contrast, vowel height both conditions overall vowel duration, with high vowels being substantially shorter than non-high vowels ($\hat{\beta}$ = -0.14, CrI = [-0.20,-0.09],

$\Pr(\hat{\beta} < 0) = 1$), and modulates the VE size. ($\hat{\beta}$ = -0.07, CrI = [-0.13,0.00], $\Pr(\hat{\beta} < 0)$ = 0.98). As Figure 1 (inner left) shows, the VE is predicted to be larger for non-high vowels than for high vowels. Both speech rate measures (mean and local) have separate effects on vowel duration (mean: $\hat{\beta}$ = -0.20, CrI = [-0.23,-0.17], $\Pr(\hat{\beta} < 0) = 1$; local: $\hat{\beta}$ = -0.28, CrI = [-0.33,-0.23], $\Pr(\hat{\beta} < 0) = 1$). There is only evidence for the effect of local speech rate on VE ($\hat{\beta}$ = -0.06, CrI = [-0.10,-0.01], $\Pr(\hat{\beta} < 0)$ = 0.99) where faster speech results in smaller VE size (shown in Figure 1, inner right), but not for mean speech rate ($\hat{\beta}$ = -0.03, CrI = [-0.09,0.02], $\Pr(\hat{\beta} < 0)$ = 0.89). This suggests that the VE is only observable for slow speech *relative* to a speaker's average speech rate, as opposed to slower-than-average speakers. Word frequency exhibits a strong effect on both vowel duration ($\hat{\beta}$ = -0.05, CrI = [-0.09,-0.01], $\Pr(\hat{\beta} < 0) = 1$), meaning that higher frequency words have shorter vowels on average. Frequency also reduces VE size ($\hat{\beta}$ = -0.07, CrI = [-0.13,-0.02], $\Pr(\hat{\beta} <) 0 = 1$), meaning that less frequent words have larger VEs. The effects of speech rate and word frequency reflect the observation that more reduced and more predictable words exhibit smaller VE values. Figure 1 (far right) illustrates the model's 95% CI for the marginal predicted effect of voicing as a function of frequency. Here, it can be observed that the model is 95% certain that the VE for sufficiently low-frequency words is larger than 1 ($\hat{\beta}$ = 0.15, CrI = [0.07,0.24], $\Pr(\hat{\beta} < 0) = 1$): that is, the size of the VE is larger for lower-frequency words, and a clear VE exists only for sufficiently low-frequency words.

4.2 RQ2: dialectal variability in VE

The dialect-level variation in VE size (between 0.08 and 0.19, median = 0.09) is roughly as large as the overall population-level VE (between 0 and 0.16, median = 0.08), though there is little evidence that the degree of between-dialect variability is *greater* than the magnitude of the population-level size of the effect ($\hat{\beta}$ = 0.04, CrI = [-0.03,0.13], $\Pr(\hat{\beta} < 0)$ = 0.82). Figure 2 shows the predicted VE sizes for each dialect: here it can be seen that dialects appear to differ *gradiently* from each other, ranging from dialects for which there is very likely no meaningful VE, to those for which there is strong evidence of a large VE size. For the Scottish dialects of Central, Glasgow, and Edinburgh, any VE size must be below 1.01, 1.03, and 1.07 respectively, based on the 95% CrI upper limit (Central: $\hat{\beta}$ = -0.05, CrI[-0.11,0.02]; Glasgow: $\hat{\beta}$ = -0.01, CrI[-0.05,0.04]; Edinburgh: $\hat{\beta}$ = -0.01, CrI = [-0.10,0.07]). This suggests that these dialects have an incredibly small—effectively null—effect of consonant voicing. For speakers from Insular and Northern Scottish regions, NYC, Lower South, and RP speakers, the range of possible VE values also includes null or negative values, where the VE value must be below e.g., 1.12 for Northern and 1.25 for NYC speakers

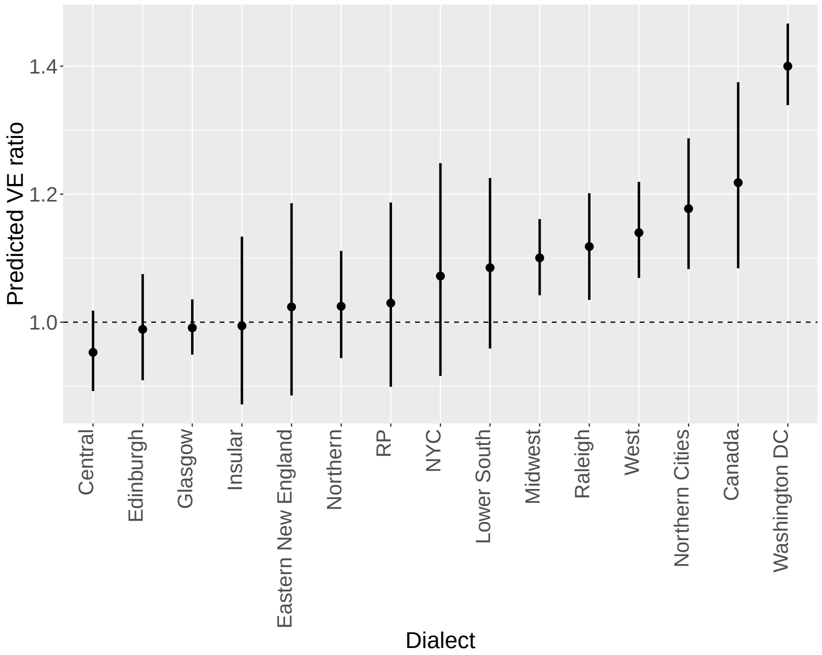

Figure 2: Estimate and 95% CrI of voicing effect for each dialect. Computed from model posterior, marginalising over all other predictors (e.g. average speech rate).

(Insular: $\hat{\beta}$ = -0.01, CrI = [-0.14,0.13]; Northern: $\hat{\beta}$ = 0.02, CrI = [-0.06,0.011]; NYC: $\hat{\beta}$ = 0.07, CrI = [-0.09,0.22], Lower South: $\hat{\beta}$ = 0.08, CrI = [-0.04,-0.20]; RP: $\hat{\beta}$ = 0.03, CrI = [-0.11,0.17]). All dialects with predicted VE values confidently above 1 (on the right side of Fig. 2) are North American varieties, though all differ in the range of possible VE values each could have. The Midwest dialect, for example, has a tight distribution of consonant voicing, where the largest possible VE value is approximately 1.16 ($\hat{\beta}$ = 0.10, CrI = [0.04,0.15]), compared with 1.20 for Raleigh ($\hat{\beta}$ = 0.11, CrI = [0.03,0.18]), 1.22 for West ($\hat{\beta}$ = 0.13, CrI = [0.06,0.20]), 1.29 for Northern Cities ($\hat{\beta}$ = 0.16, CrI = [0.08,0.22]), and 1.37 for Canada ($\hat{\beta}$ = 0.20, CrI = [0.08,0.32]). AE speakers from Washington DC exhibited the largest VE value in the sample of dialects, with a median VE of approximately 1.40 and a maximum possible VE of 1.47 ($\hat{\beta}$ = 0.34, CrI = [0.29,0.38]), similar to elicited AAE speech in North Carolina (Holt et al., 2016).

5 Discussion

The VE is a well-documented property of English and a number of other languages, in which vowels preceding voiced consonants are longer than their voiceless counterparts. Languages vary in both the size and presence of the effect (Keating, 1985), it has been shown to be subject to speaking style effects (Port, 1977), and it is considered to be at least in part under the control of speakers (Solé, 2007). The size of this effect is considered to be larger for English than in other languages (Chen, 1970), and has been shown to play a significant effect in the perception of the final consonant's voicing specification (Denes, 1955; Klatt, 1976). Given that much of the prior research on the English VE has focused on production in elicited laboratory settings (House and Fairbanks, 1953; House, 1961; Holt et al., 2016), far less is known about how variable the VE is in spontaneous connected speech. As vowel duration has been shown to be substantially affected by a range of linguistic factors (such as vowel height, speech rate, consonant manner, word frequency), it is unclear as to the extent by which VE is also separately modulated by these factors outside of laboratory speech. Furthermore, the vast majority of studies on English VE have focused on North American English, and only a handful of studies have empirically examined dialectal variability (Hewlett et al., 1999; Jacewicz et al., 2007; Tauberer and Evanini, 2009; Holt et al., 2016). As multiple dialect regions have not been examined in a single instance, simultaneously controlling for linguistic factors on vowel duration in the same way, it is not clear whether the observed differences in VE represent discrete phonological implementations of the effect, or rather a continuum of possible values for English VE.

In this study, a weak effect of consonant voicing on vowel duration was observed, looking at connected speech from 15 dialects of North American and British English speech. This effect was weaker than previously observed in laboratory studies of VE. The VE was also shown to be modulated by a range of linguistic factors. Specifically, the size of the VE was affected by consonant manner, vowel height, speech rate, and lexical frequency. This suggests that the VE is stronger for non-high vowels, before stops, for less reduced vowels, and for less frequent words. Whilst much of the original phonetic literature on VE (House and Fairbanks, 1953; House, 1961) used data produced as single words in slow speech and controlled phonetic conditions, this study demonstrates both that VE is sensitive to both speech style and to its phonetic and phonological environment.

The VE was also shown to be highly variable across dialects, with some dialects exhibiting effectively null VE values, to dialects with large and robustly positive VEs. The white AE dialects with confidently-positive VEs (Midwest, Raleigh, West, Northern Cities, Canada) have values broadly similar to those observed in spontaneous speech by Tauberer and Evanini (2009). This similarity provides further evidence that speech style (i.e., spontaneous connected speech versus lab-elicited speech) plays a substantial role on the production of VE, and that controlling for the influence of other linguistic factors (e.g., speech rate, frequency, etc.) does not adversely affect the overall VE size for these dialects. This range of VE values across all dialects studied here is not inconsistent with the possibility that dialects can be clustered into groups with and without robust VEs; a number of dialects exhibit a wide range of potential VE values (as reflected in the large credible intervals in the model), and so it is not obvious how to characterise the differences between dialects. Dialects being distinguished into a binary classification of either 'having' or 'not having' a strong VE would not be unlike the observation made of language-level distinctions between VE-full languages such as English, French, Russian (Chen, 1970), and VE-less languages such as Polish (Keating, 1985) or Japanese (Port et al., 1987; Han, 1994). At the same time, these findings are also not inconsistent with dialects operating across a spectrum of VE sizes, wherein dialects differ gradiently from each other. In contrast to a binary 'presence or absence' of VE, dialects and languages may be distinguished in the degree of VE. As the previous studies of VE across languages were performed using laboratory speech, it would be worth understanding how the VE is realised in spontaneous speech of languages other than English—if other languages (such as French, Korean, etc.) exhibit a smaller VE in citation speech (as Chen (1970) suggests), it is not clear how the VE would present itself in less-controlled linguistic environments.

One point to note is that the consonant voicing in this study is defined phonologically: the particular voicing of a consonant is determined by the phone label for the consonant used during forced alignment (Buckeye is an exception to this, where phones were transcribed phonetically). As a result this study does not examine the production of VE based on the *phonetic* realisation of voicing for the vast majority of corpora. If VE is (in part) phonetically-driven (i.e., as an articulatory consequence of the production of voicing), analysing VE as a function of *phonological* voicing may have some substantial consequences, such as not accounting for whether a stop has been realised as a glottal stop or has been phonetically deleted. This has potential consequences for both British Englishes and AAE: the neutralisation of final consonant voicing is considered to be the trigger for the large VE in AAE (Adams, 2009; Holt et al., 2016), whilst glottalisation of voiceless stops is an integral feature of British Englishes (Kerswill, 2003). It is possible, then, that the particular phonetic realisation of stops may also be relevant as how the VE is realised in particular dialects.

6 Conclusion

This study examined how the durational difference between vowels preceding voiced or voiceless consonants varied in a range of phonological, phonetic, and dialectal contexts. In comparison with previous laboratory studies, this difference was smaller in spontaneous connected speech, with larger differences observed in slow speech, before stops, for high vowels, and for low frequency words. Dialects varied substantially in the size of this effect, ranging from dialects with a negligible difference to those with almost

50% difference between pre-voiceless and pre-voiced vowels. These results suggest that the size of the voicing effect is more subtle in spontaneous connected speech, and that a range of dialect-specific sizes are observable, in contrast to there being a single 'English' voicing effect.

Acknowledgements 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). The SPADE project team would like to thank Anne Fabricius, the Data Guardian of the Modern RP corpus, for allowing this corpus to be used in developing the ISCAN software.

References

Adams, C. A. (2009). An acoustic phonetic analysis of African American English: a comparative study of two dialects. Master's thesis, Eastern Michigan University.

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 Beal, J. C., Corrigan, K. P., and Moisl, H. L., editors, *Creating and Digitizing Language Corpora*, pages 17–34. Palgrave, New York.

Ayelett, M. and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, 119:3048–3058.

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.

Brysbaert, M. and New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods*, 41:977–990.

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1):395–411.

Bybee, J. B. (2001). *Phonology and Language Use*. Cambridge University Press, Cambridge.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22:129–159.

Crystal, T. H. and House, A. S. (1982). Segmental durations in connected speech signals: preliminary results. *Journal of the Acoustical Society of America*, 72:705–716.

Crystal, T. H. and House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, 88:101–112.

Cuartero, N. (2002). *Voicing assimilation in Catalan and English*. PhD thesis, Universitat Autònoma de Barcelona.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, 27:761–764.

Dodsworth, R. (2013). Retreat from the Southern Vowel Shift in Raleigh, NC: social factors. *University of Pennsylvania Working Papers in Linguistics*, 19:31–40.

Dodsworth, R. and Kohn, M. (2012). Urban rejection of the vernacular: The SVS undone. *Language Variation and Change*, 24:221–245.

Ernestus, M. and Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39:253–260.

Fabricius, A. H. (2000). *T-glottalling between stigma and prestige: a sociolinguistic study of Modern RP*. PhD thesis, Copenhagen Business School, Copenhagen, Denmark.

Farrington, C. (2018). Incomplete neutralization in African American English: the cast of final consonant devoicing. *Language Variation and Change*, 30:361–383.

Fromkin, V. A. (1977). Some questions regarding universal phonetics and phonetic representations. In Juilland, A., editor, *Linguistic studies offered to Joseph Greenberg on the occasion of his sixtieth birthday*, pages 365–380. Anma Libri, Saratoga.

Fruehwald, J. (2013). *The Phonological Influence on Phonetic Change*. PhD thesis, University of Pennsylvania.

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

Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, Cambridge.

Greenbaum, S. and Nelson, G. (1996). The International Corpus of English (ICE project). *World Englishes*, 15:3–15.

Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society* *of America*, 96:73–82.

Harris, M. and Umeda, N. (1974). Effect of speaking mode on temporal factors in speech: vowel duration. *The Journal of the Acoustical Society of America*, 56(3):1016–1018.

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.

Holt, Y. F., Jacewicz, E., and Fox, R. A. (2016). Temporal variation in African American English: the distinctive use of vowel duration. *Journal of Phonetics & Audiology*, 2.

House, A. S. (1961). On vowel duration in English. *Journal of the Acoustical Society of America*, 33:1174–1178.

House, A. S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 25:105–113.

Jacewicz, E., Fox, R. A., and Salmons, J. (2007). Vowel duration in three American English dialects. *American Speech*, 82:367–385.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (2001). Probabilistic relations between words: evidence from reduction in lexical production. In Bybee, J., editor, *Frequency and the Emergence of Linguistic Structure*, pages 229–254. John Benjamins, Amsterdam.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In Fromkin, V. A., editor, *Phonetic Linguistics: essays in honor of Peter Ladefoged*, pages 115–132. Academic Press, New York.

Kendall, T. and Farrington, C. (2018). The Corpus of Regional African American Language. Version 2018.10.06.

Kerswill, P. (2003). *Dialect levelling and geographical diffusion in British English*, pages 223–243. John Benjamins, Amsterdam.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America*, 54:1102–1104.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5):1208–1221.

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

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100:1989–2001.

Luce, P. A. and Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America*, 78:1949–1957.

McAuliffe, M., Coles, A., Goodale, M., Mihuc, S., Wagner, M., Stuart-Smith, J., and Sonderegger, M. (2019). ISCAN: A system for integrated phonetic analyses across speech corpora. In *Proceedings of the 19th Congress of Phonetic Sciences (ICPhS2019)*, Melbourne.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage, P. F., editor, *The Production of Speech*, pages 189–216. Springer, New York.

Paape, D., Nicenboim, B., and Vasishth, S. (2017). Does antecedent complexity affect ellipsis processing? an empirical investigation. *Glossa*, 77:1–29.

Peterson, G. E. and Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32:693–703.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., and Fosler-Lussier, E. (2007). *Buckeye Corpus of Spontaneous Speech*. Ohio State University, Columbus, 2 edition.

Port, R. (1977). *The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words*. PhD thesis, University of Connecticut.

Port, R. F., Delby, J., and O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, 81:1574–1585.

Priva, U. C. (2017). Informativity and the actuation of lenition. *Language*, 93:569–597.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, 51:1296–1303.

Rathcke, T. and Stuart-Smith, J. (2016). On the tail of the Scottish Vowel Length Rule in Glasgow. *Language and Speech*, 59:404–430.

Solé, M.-J. (2007). Controlled and mechanical properties in speech. In Beddor, P. and Ohala, M., editors, *Experimental Approaches to Phonology*, pages 302–321. Oxford University Press, Oxford.

Solé, M.-J. and Ohala, J. J. (2010). What is and what is not under the control of the speaker: intrinsic vowel duration. *Laboratory Phonology*, 10:607–655.

Stan Development Team (2018). RStan: the R interface to Stan. R package version 2.18.2.

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 Montgomery, C. and Moore, E., editors, *Language and a Sense of Place: Studies in Language and Region*, pages 38–65. Cambridge University Press, Cambridge.

Summers, W. V. (1987). Effects of stress and final consonant voicing on vowel production: articulatory and acoustic analyses. *Journal of the Acoustical Society of America*, 82:847–863.

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 2009*.

Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, 58:434–445.

Vasishth, S., Nicenboim, B., and Beckman, M. (2018). *Bayesian Data Analysis in the Phonetic Sciences: A Tutorial Introduction*. OSF.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91:1707–1717.

Wilson, C. and Chodroff, E. (2017). Uniformity of intrinsic vowel duration across speakers of American English. In *The 174th Meeting of the Acoustical Society of America*, New Orleans.