

Structured heterogeneity in Scottish stops over the twentieth century

Morgan Sonderegger^{*1}, Jane Stuart-Smith^{†2}, Thea Knowles^{‡3}, Rachel MacDonald^{§2}, and Tamara Rathcke^{¶4}

¹Department of Linguistics, McGill University

¹Centre for Research on Brain, Language, and Music

²University of Glasgow

³University of Buffalo

⁴University of Kent

Preprint version of September 2019, to appear in *Language*

Abstract

How and why speakers differ in the phonetic implementation of phonological contrasts, and the relationship of this ‘structured heterogeneity’ to language change, has been a key focus over 50 years of variationist sociolinguistics. In phonetics, interest has recently grown in uncovering ‘structured variability’—how speakers can differ greatly in phonetic realization in non-random ways—as part of the longstanding goal of understanding variability in speech. The English stop voicing contrast, which combines extensive phonetic variability with phonological stability, provides an ideal setting for an approach to understanding structured variation in the sounds of a community’s language which illuminates both synchrony and diachrony. This paper examines the voicing contrast in a vernacular dialect (Glasgow Scots) in spontaneous speech, focusing on individual speaker variability within and across cues, including over time. Speakers differ greatly in the use of each of three phonetic cues to the contrast, while reliably using each one to differentiate voiced and voiceless stops. Inter-speaker variability is highly structured: speakers lie along a continuum of use of each cue, as well as correlated use of two cues—VOT and closure voicing—along a single axis. Diachronic change occurs along this axis, towards a more aspiration-based and less voicing-based phonetic realization of the contrast, suggesting an important connection between synchronic and diachronic speaker variation. **Keywords:** phonetic variation, sound change, sociolinguistics, stop voicing, individual differences, Scottish English *

^{*}morgan.sonderegger@mcgill.ca

[†]Jane.Stuart-Smith@glasgow.ac.uk

[‡]theaknow@buffalo.edu

[§]Rachel.Macdonald@glasgow.ac.uk

[¶]T.V.Rathcke@kent.ac.uk

^{*}We thank three anonymous reviewers, Megan Crowhurst, and Khalil Iskarous for helpful comments on this

1 Introduction

The recognition of observable structured linguistic variation, *structured heterogeneity*, as an essential fact of language, inherently providing order to linguistic systems at any one time, and the impetus for language change over multiple time points, was first advanced by Weinreich et al. (1968) to heal the rift between historical and synchronic linguistics. Weinreich et al. argued for a refocusing on structured variation within and across speakers in their communities, constrained by linguistic and social factors, as the locus to seek empirical evidence for language change. While their manifesto was explicitly directed towards enabling explanations for language change, it was also a more general call to integrate synchronic and diachronic approaches to language study by focusing on empirical observation of structured heterogeneity (Weinreich et al., 1968, p. 101).

The English stop voicing contrast provides an ideal phenomenon for returning to Weinreich et al.’s vision, in which the description of structured heterogeneity in a community’s language gives insights for synchrony and diachrony together (Purnell et al., 2005). The West Germanic stop voicing contrast has been largely stable at a phonological level for a long time, yet its phonetic realization is variable both within and across time periods and languages/varieties (e.g. Honeybone, 2005; Purnell et al., 2005; Salmons, 2019). Research on production of the English stop voicing contrast, largely through controlled phonetic studies of read speech, has found that many acoustic cues signal the contrast (Klatt, 1975; Lisker, 1986), especially Voice Onset Time (VOT), but also voicing during closure, closure duration, and others (e.g. Abramson & Whalen, 2017; Byrd, 1993; Davidson, 2016, 2017; Lisker & Abramson, 1964). And, as Purnell et al. (2005) observe in their study of changes to the Wisconsin English word-final stop contrast, the ‘trading relations’ (Repp, 1982) inherent between these cues to laryngeal timing provide an effective vehicle which is exploited for change to take place.

Recent phonetic studies have begun to focus on a specific aspect of stop variation. ‘Structured variability’ (Chodroff et al., 2015) refers to the set of observations that individual speakers may differ from each other in the nature and range of their phonetic variation, in ways that are not random. Speakers can differ significantly from each other in the use of one or more phonetic cues in ways that are in part explicable (e.g. by speaking style: Clayards, 2018); they can show covariation in the use of a single cue to signal linguistic contrasts across different phonological categories (Chodroff & Wilson, 2017; Theodore et al., 2009); or covariation in the use of multiple cues to signal the same linguistic contrast (Bang, 2017; Clayards, 2018; Shultz et al., 2012). Relatively little of the possible structured variability in the English stop voicing contrast has been mapped out: how and why individual speakers vary in their realization of one or more cues for single stop categories, or the stop voicing contrast itself, is still far from clear. Phonetic studies in this area have also largely not considered spontaneous speech, where individual variability in a phonetic cue’s realization must be disentangled from the many linguistic and prosodic factors (e.g. place of articulation, speech rate) which affect cues in natural speech.

This paper offers an integrated account of an English stop voicing contrast which shows

paper; as well as Joseph Salmons, Meghan Clayards, James Tanner, and audiences at NWAV 2015 and LANCHART (University of Copenhagen, September 2015) for feedback on this project. We thank Cordelia Klein and Misha Schwartz for research assistance. Construction and preliminary analysis of the Sounds of the City corpus were funded by the Leverhulme Trust (#RPG-142). Work reported in this paper was supported by the Social Sciences and Research Council of Canada (#430-2014-00018, #435-2017-0925) and the Fonds de Recherche du Québec Société et Culture (#183356);

that speakers exhibit structured variability in multiple cues, and that a subspace of this variability acts as the multidimensional axis along which diachronic change is progressing, thus exploiting the ‘trading relations’ between the cues to voicing in the change. We examine English stop voicing in a specific sociolinguistic context—spontaneous Glasgow Scots vernacular over the course of the twentieth century—which is interesting and informative for three reasons:

1. **Phonetic realization:** Scots shows a more recessive version of the English stop voicing contrast, with relatively more voicing in voiced stops and shorter lag voiceless stops (e.g. Scobbie 2006). Few studies have considered individual speaker variability in the Scottish English stops, and none have considered more than one cue.
2. **Speech style:** Naturally-occurring spontaneous speech data is available. Almost all previous studies of the English stop voicing contrast have been on different forms of read speech, and have also given hints that phonetic cues to voicing are enhanced in controlled speech. Our study is the first to systematically examine structured variability for English stop voicing in casual speech.
3. **Time:** Previous work suggests a shift in Scots from a more voicing-based stop contrast to a more aspiration-based contrast over the course of the twentieth century. Our data are drawn from a real-time corpus of Glasgow vernacular with an effective real- and apparent-time span of around 100 years. Most studies of English stop voicing have considered recordings from speakers of the same age made at the same time point, and no previous studies have examined multiple cues to the voicing contrast by sampling spontaneous speech over real time.

We analyse three temporal phonetic cues to the Glaswegian stop voicing contrast—(positive) Voice Onset Time, the degree of voicing during closure, and closure duration—with particular focus on the ways in which individual speakers vary within and across cues, including over time.

1. Do individual speakers differ in their use of single phonetic cues to realize Scottish stop voicing?
2. How do individual speakers coordinate multiple phonetic cues to realize this contrast?
3. Is the use of individual speakers’ cues also structured by decade of recording, suggesting change in phonetic realization in Scottish stops over time?

We take structured variability in the realization of the stop voicing contrast as a specific instantiation of structured heterogeneity. We define phonetic variation to be **STRUCTURED** if we observe individual differences in how speakers use one or more cues to realize the voicing contrast which are demonstrably non-random: speaker differences in how a single cue is used to signal both voiced and voiceless categories or the contrast between them (RQ1); covariation across speakers in how they use multiple cues (RQ2); or individual differences which are explained by the social factor of time for this community (RQ3).

The paper is organized as follows. After considering relevant literature (Section 2), we describe our speaker sample and semi-automated methods for measuring the three acoustic cues (Section 3) to stop voicing. We then identify individual speaker variation for each cue, and across multiple cues, treating the dataset as a quasi-synchronic sample. Section 4 shows that all speakers use each cue to realize the voicing contrast, above and beyond key linguistic and prosodic factors, but with significant interspeaker variation (RQ1). Section 5

finds that these speaker differences are related: the two cues used most to realize the voicing contrast, (positive) VOT and voicing during closure (RQ2), are correlated across individual speakers. Section 6 uncovers a small but real shift in the trading relation between the cues, towards a more aspiration-based realization for the stop voicing contrast (RQ3)—exactly the multidimensional axis along which interspeaker variability was observed, synchronically.

2 Background

2.1 Cues to English stop voicing

The English stop system continues that of West Germanic (Salmons, 2019). Phonologically there are two series, referred to here as ‘voiceless’ and ‘voiced’ (phonemic /p t k/, /b d g/), differing in one phonological feature. There is debate whether that feature is [voice] or [spread] glottis (see Honeybone, 2005). The ‘laryngeal realism’ view (e.g. Iverson & Salmons, 1999; Salmons, 2019), in favor of [spread], is motivated both by phonological processes and by phonetic realization of the two stop series in many English dialects as voiceless aspirated and voiceless unaspirated stops [p^h t^h k^h] and [p t k], especially in stressed syllable-initial position.

Within these phonological categories, the phonetic realization of voiced and voiceless stops varies by dialect and time (Section 2.5). While VOT is the most widely studied, there are many other spectral and temporal cues to the voicing contrast (Lisker, 1986). We review the findings for the three temporal cues analysed here: VOT, voicing during closure, and closure duration. All studies referred to examine British English (Docherty, 1992; Sonderegger et al., 2017; Stuart-Smith et al., 2015; Summerfield, 1975) or American English (all others).

VOICE ONSET TIME (VOT), the time between the stop release and the initiation of glottal pulsing, has been shown to be the most important cue to the English stop voicing contrast, especially in word-initial position, from many studies on read standard varieties of English (beginning with Liberman et al., 1958; Lisker & Abramson, 1964), and a few studies of spontaneous English (Baran et al., 1977; Sonderegger et al., 2017; Stuart-Smith et al., 2015; Yao, 2009). Voiceless stops show positive VOT values (long lag), whereas voiced stops show either shorter positive VOT values (short lag), and/or negative VOT (voicing lead), reflecting the presence of voicing during closure and before stop release. VOT is also affected by linguistic and prosodic constraints, including place of articulation (Docherty, 1992; Lisker & Abramson, 1964), speech rate (differently for voiced/voiceless stops: e.g. Kessinger & Blumstein, 1997; Summerfield, 1975), phrasal accent (Cole et al., 2007), and position in the phrase (Yao, 2009).

English stop VOICING DURING CLOSURE, the duration of glottal pulsing between stop closure onset and the burst, has to our knowledge been studied exclusively in read speech. Davidson (2016, 2017) provides the most recent detailed phonetic examination, for American English stops. For voiced stops, full voicing during closure is more likely in labial than velar stops (Docherty, 1992), in phrase-medial/final position (Docherty, 1992; Lisker & Abramson, 1964, 1967), in word-medial position (cf. Keating, 1984), when closure duration is shorter (cf. Westbury & Keating, 1986), and when the preceding sound is (phonologically) voiced (cf. Docherty, 1992; Lisker & Abramson, 1964). Voicing in voiceless stops follows similar constraints, occurring more in phrase-medial/final position, in word-medial position, and following a vowel or approximant (Docherty, 1992). ‘Negative VOT’ is often observed for voiced stops in phrase-initial position (e.g. Keating, 1984; Lisker & Abramson, 1964, 1967),

but there is no established way to extend this concept to phrase-medial stops, where it overlaps with ‘voicing during closure’. A common operationalization is that VOT is ‘negative’ if voicing begins after the onset of closure and continues past the release; a more recent proposal is that VOT be defined as negative when the closure is >50% voiced (Abramson & Whalen, 2017). Using the first definition, Davidson (2016) found negative VOT in only 1% of voiced stops.

English stop CLOSURE DURATION has also been examined largely in read speech (except Yao, 2007, on spontaneous speech). Several studies have found shorter closure duration for voiced stops (e.g. Chen, 1970; Luce & Charles-Luce, 1985), though not all (Crystal & House, 1988); Byrd (1993) finds the voiced/voiceless difference is modulated by place of articulation (1–6 ms for different places). Zue (1976) and Yao (2007) consider voiceless stops alone; Yao finds effects of place of articulation, preceding phone, and speech rate.

2.2 Individual speaker variation for single cues to English stop voicing

Early work on English VOT already noted that individual speakers vary systematically in how VOT is used to signal the voicing contrast (Lisker & Abramson, 1964, p. 395).¹ Allen et al. (2003) established individual speaker differences in VOT in read monosyllables, which remained after controlling for an individual’s speech rate. Theodore et al. (2009) replicated this finding, and showed that individual speakers’ VOT values for /p/ and /k/ were tightly correlated, such that each speaker distinguished place of articulation.

Chodroff & Wilson (2017) extended this finding of structured variability to less controlled speech, in their analysis of positive VOT in voiced and voiceless stops, from 180 speakers of different ages and American English dialects, in two speech styles (read monosyllables, read sentences). Speakers’ VOT values for nearly all pairs of stops with the same voicing (e.g. /p/~/t/, /g/~/b/) were positively correlated in both speech styles, with the result that each speaker reliably distinguished place of articulation. Similarly, individual speakers tended to distinguish voiced and voiceless stops (e.g. positive /p/~/b/ correlation)—more clearly in read sentences than in isolated words. The same speakers showed structured variability for two additional cues, following vowel fundamental frequency at onset (here, ‘onset f0’) and burst spectral center (Chodroff & Wilson, 2018). While Chodroff & Wilson (2017, 44–45) report a preliminary extension to spontaneous speech, whether similar structured speaker variation holds in spontaneous speech—or for voicing and closure duration—is unknown.

2.3 Individual speaker variation across multiple cues to English stop voicing

Perceptual studies on multiple cues to English stop voicing have mainly examined VOT and onset f0 in word-initial stops. VOT always emerges as the primary cue, though its relationship with onset f0 varies according to the voicing of the stop and the variety of English (e.g. Francis et al., 2008; Schertz et al., 2015). Less is known about PRODUCTION of multiple cues to English stop voicing by individual speakers.

The few recent studies suggest that, as for single cues, individual speakers can systematically differ from each other in how they use multiple cues to contrast English stops. Both

¹We do not consider here individual speaker variation in phonetic realization, including change over time, for stop contrasts in languages other than English.

Shultz et al. (2012) (read monosyllables) and Bang (2017) (read sentences) found significant correlations for VOT and onset f_0 , showing that the more a speaker used VOT to cue the voicing contrast, the less they used f_0 , suggesting a trade-off at the level of individual speakers. Schertz et al. (2015) considered the relative contribution of VOT, onset f_0 , and closure duration in read minimal pairs for L2 Korean English speakers. Individual speakers did not all show the same pattern of cue use to mark the contrast, perhaps because it is a non-native variety. Clayards (2018) also analysed three cues to stop voicing—VOT, onset f_0 , and relative vowel duration following the stop—in American English speakers reading minimal pairs. She found systematic talker differences in use of the three cues for each stop, consistent with a strong effect of speech style—hyperarticulation from reading minimal pairs—which seemed to induce prototypical stop productions for some speakers.

These studies show both that speakers vary in the coordination of multiple cues to signal stop contrasts, and that structures are present within this variation. The focus to date has been on structure relating to the maintenance of linguistic contrasts in highly controlled speech styles. Reasons for why speakers might differ from each other, while retaining linguistic contrasts, are not explored, though speaker dialect is sometimes mentioned and speech style seems likely for Clayards (2018). Chodroff & Wilson (2018) point to diachronic variation as a possible factor, though there has been little consideration of sound change in multiple cues to English stop voicing. To our knowledge there is only Purnell et al. (2005)’s qualitative analysis of voicing in Wisconsin English stops and fricatives, which shows shifts in individual speakers’ relative use of voicing during closure and preceding-vowel duration, exploiting the inherent trading relation between these cues, across a real- and apparent-time span of over 100 years.

2.4 Speech style and stop voicing

It has been long noted that speech style affects English stop voicing, even in read speech. Lisker & Abramson (1967) examined speech style as one contextual factor affecting VOT. In isolated words, stops had longer VOTs, but there was even better separation of voiced and voiceless stops in minimal pairs (‘enhancement’), compared to read sentences. Similarly, Baran et al. (1977) found the voiced/voiceless VOT difference was reduced in conversational versus citation-form speech for three speakers, while Chodroff & Wilson (2017) found shorter VOT means in read sentences than in monosyllables. Chodroff & Wilson were also struck by the persistence of strong interspeaker co-variation patterns in connected speech where many more factors influence VOT. Our study of cue co-variation in spontaneous speech permits examination of whether speakers maintain contrasts in a much less controlled speech style.

2.5 Changing stops in Scottish English

Scottish English is a sociolinguistic continuum, from broad Scots, continuing forms of Northern Anglian, spoken by working-class speakers, to Scottish Standard English, continuing early 17th century Southern English, spoken by middle-class speakers (e.g. Aitken, 1979). Scottish Standard English voiceless stops have been reported to be phonetically less aspirated than in Southern British Standard English at least since the turn of the twentieth century (Grant, 1913, p. 80, Wells, 1982). Masuya (1997)’s small-scale study shows shorter VOT in Standard Scottish English than in Anglo-English. Pre-Second World War descriptions of Scots also report unaspirated syllable-initial voiceless stops (Johnston, 1997, p. 505). Scobbie (2006)’s

study of speakers in Shetland demonstrates the social-dialectal range, with less/more aspirated /p/ and more/less voiced /b/, depending on Shetlandic/mainland-Scottish heritage.

Phonologically, there is debate whether the ‘more voiced’ nature of stop realization in traditional Scots reflects a difference in phonological representation from most dialects of English. Iverson & Salmons (1999) and Salmons (2019) argue, based on both phonetic realization and phonological processes, that under ‘laryngeal realism’ the voicing contrast in traditional Scots is in [voice], in contrast to [spread] in other English dialects. The traditional view would be that Scots and all other English varieties have a [voice] contrast, with dialect-specific phonetic implementation.

Two changes are claimed to be taking place at either end of the Scottish English continuum. First, Scottish Standard English stops may be becoming more aspirated, as in Anglo-English varieties (Masuya, 1997). Docherty et al. (2011)’s study of VOT from read wordlists in 159 speakers from the Scottish-English Border also found longer VOT and less frequent use of negative VOT in younger speakers, with further patterning according to affiliation with Scottish identity. Second, stops in Scots may be becoming more aspirated, from leveling towards Scottish Standard English (Johnston, 1980, p. 78 in Scobbie, 2006, p. 374). Stuart-Smith et al. (2015) found significantly longer positive VOT in a real- and apparent-time study of stops in spontaneous Glaswegian Scots, with a twist: the youngest speakers, born in the 1990s, showed a reversal, with short VOTs matching those of much older/earlier-born speakers.

Thus it appears that the Scots stop voicing contrast is shifting from a recessive, more voicing-based system, to a more aspiration-based system. But previous studies of Scottish English have been limited to a single cue (VOT), largely restricted to read speech, and have mostly focused on the group rather than individual speakers. We do not know what degree of structured variability may exist for individual speakers across stop voicing cues in spontaneous speech. We address these issues here with an examination of the Scots end of the Glaswegian Scottish English continuum.

3 Methodology

We analyzed three temporal acoustic cues to the Glaswegian stop voicing contrast in stressed syllable-initial stops: positive VOT (VOT), degree of Voicing During Closure (VDC), and Closure Duration (CD). (We capitalize all three cues, for consistency.) We first describe the speech corpus used, then turn to the datasets of the three acoustic cues, each of which was built ‘semi-automatically’, through automatic measurement followed by manual correction.

3.1 Sample

We examined speech from the SOUNDS OF THE CITY corpus of Glasgow vernacular (e.g. Stuart-Smith et al. 2017; Stuart-Smith & Lawson 2017). This corpus consists of audio recordings and orthographic transcripts, stored and force-aligned at the segment level using LaBB-CAT (Fromont & Hay, 2012), from over 140 working-class speakers. The corpus is structured by decade of recording (1970s, 1980s, 1990s and 2000s) and by speaker age (older: aged 67–90; middle: aged 40–55; young: aged 10–17),² which allows investigation of sound change across the twentieth century in real and apparent time.

²The larger gap between young and middle-aged speaker groups was partly the result of the recordings available, and partly to ensure clear generational separation between adolescents and their parents.

Table 1: The sample of 23 female speakers from the SOUNDS OF THE CITY corpus analysed in this study. Decade of birth of each group of speakers is given in parentheses. 3 (1890s) = ‘3 female speakers born in the 1890s’, and so on.

DECADE OF RECORDING	AGE		
	Elderly (67–90)	Middle-aged (40–55)	Young (10–17)
1970s	3 (1890s)	4 (1920s-b)	4 (1960s)
2000s	4 (1920s-a)	4 (1950s)	4 (1990s)

We use the same subset of the corpus as in Stuart-Smith et al. (2015) (Table 1): 23 female speakers, from the three age categories, recorded in the 1970s and the 2000s. We work with female speakers for two reasons. First, a primary aim of this line of work is assessing change over time in the Scottish English voicing contrast—our third research question—and to have a large enough sample size to examine the effect of time while holding other social factors constant, it was necessary to restrict to speakers of one gender. Second, given that changes from below are often led by female speakers (e.g. Labov 2001), we assumed that any indications of change would be most evident for this gender.

3.2 Voice Onset Time (VOT)

The VOT dataset is the same as that used in Stuart-Smith et al. (2015), where full details of the VOT measurement methodology are given. In summary, an automatic measurement of positive VOT was obtained for all stressed, syllable-initial stops using AutoVOT (Keshet et al., 2014; Sonderegger & Keshet, 2012) followed by manual correction. Tokens where the automatic measurement was correct or which could be easily corrected were kept in the dataset, and the remaining 25.9% tokens were discarded (e.g. realization as a fricative, approximant, or glottal; gross forced-alignment errors). The final dataset contained 4096 voiced tokens and 3254 voiceless tokens ($n = 7350$ total).

Stops fitting the traditional definition of ‘negative VOT’—where voicing began during the closure and continued to the burst—were very rare (~ 15 instances; cf. Davidson 2016), unlike previous studies of Scottish English based on read speech (Docherty et al., 2011; Masuya, 1997; Scobbie, 2006). Voicing during stop closure instead tended to appear as continuous throughout the entire closure, as no voicing at all, or as ‘bleed’ voicing from a preceding voiced segment for a fraction of the closure (Davidson, 2016). Rather than re-defining VOT in a way that would give these tokens negative values (e.g. Abramson & Whalen, 2017), we defined VOT to be a strictly positive measure—similarly to Davidson (2016) and Kim et al. (2018)—because this gave us greater flexibility to describe laryngeal timing for the Scottish stop contrast. We could measure (positive) VOT for all stops, and have this be an independent dimension to ‘voicing’, which we examined through the degree of voicing during closure (see Cho et al., 2019).

3.3 Voicing During Closure (VDC) and Closure Duration (CD)

VDC and CD were also detected using a semi-automatic process, of automatic measurement followed by manual correction. Unlike AutoVOT for VOT measurement, no specialized algorithm exists for VDC/CD measurement. Thus, semi-automatic measurement for VDC and

CD was more time consuming compared to VOT, and annotating the entire set of stops annotated for VOT was not feasible. We therefore limited the sample for VDC and CD to a subset of stressed syllable-initial stops: those which had a valid VOT measurement; were phrase-medial following a vowel or a fricative; and had automatically-measured closure duration of at least 30 ms. These two preceding environments were selected to give a range of phonetic context effects on stop voicing (vowel/voiceless fricative \Rightarrow more/less VDC; e.g. Davidson, 2016, 2017; Iverson & Salmons, 1995). Stop closures less than 30 ms were problematic for automatic voicing measurement.

3.3.1 VDC/CD Step 1: Automatic measurement

Stop closure durations were automatically measured as the interval between closure onset, the force-aligned left boundary of the stop associated with the labelled VOT, and closure offset, the onset of the VOT interval. The amount of voicing during the closure was automatically measured using a custom Praat script (Boersma & Weenink, 2001). For each stop token, the script extracted the full stop segment plus a margin of 500ms on either side, which was found to give voicing detection closer to human annotators' judgments. Voicing was extracted in this region by detecting a pitch track (cross-correlation method), inferring a point process of glottal pulses using this pitch track and the audio, then defining VDC as the interval between the last pulse and closure onset. This procedure resulted in three qualitative types of Voicing During Closure pattern: no closure voicing, full closure voicing, and perseverative voicing. Other patterns were extremely rare (e.g. 'negative VOT', noted above) and were not allowed as automatic predictions. Because 'no' and 'full-closure' voicing were so common relative to perseverative voicing, we realized early on that the analysis would only use a three-way division of VDC (None/Some/All), and manual correction corrected VDC in a way that maximized the number of tokens for analysis using this division.

3.3.2 VDC/CD Step 2: Manual Correction

Manual inspection, correction, and coding of the predicted closure and voicing boundaries were carried out by two annotators, in Praat. Closure boundaries were assumed to be fixed (from forced alignment and VOT measurement), while the right boundary of voicing could be adjusted. When the left closure boundary was inaccurate, the VDC measurement was deemed 'correctable' if correcting the closure boundary would not change whether the VDC annotation was None, Some, or All. Since the closure boundary error was often small enough to not change the VDC annotation, this method allowed many tokens with incorrect CD left boundary (and hence excluded from the CD analysis) to remain in the VDC analysis. Thus, tokens received either a CD and VDC measurement (both accurate or correctable), a VDC measurement alone (VDC correctable despite incorrect closure boundaries), or neither CD nor VDC measurements—making the CD dataset a subset of the VDC dataset. A representative Praat TextGrid with manually-corrected annotations for VOT, CD and VDC is shown in Figure 1.

The original sample of correct/corrected VOT predictions consisted of 7350 tokens. Reducing the sample to stops following vowels or fricatives, and with closure duration above 30 ms, left 4841 tokens (2715/2126 voiced/voiceless) for which VDC predictions were corrected. Of these, 1593 (844/749) did not have valid or correctable voicing intervals; the remaining 1871 voiced and 1377 voiceless tokens made up the dataset used to model VDC ($n = 3248$). Of the original 4841 tokens, 3098 (1701/1397 voiced/voiceless) had invalid closure boundary

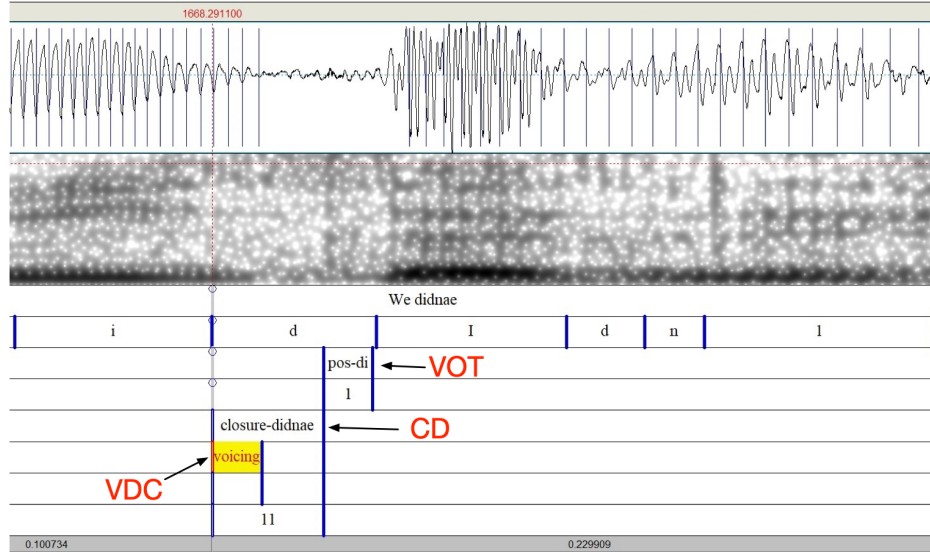


Figure 1: Part of Praat TextGrid showing VOT, CD, and VDC measurements for /d/ in “didnae”.

or boundaries; the remaining 1014 voiced and 729 voiceless tokens made up the dataset used to model CD ($n = 1743$).

4 Analysis 1: Individual speaker variation within phonetic cues

4.1 Preliminaries

For each acoustic cue (VOT, VDC, CD), the goal of Analysis 1 was to determine each speaker’s characteristic values, after controlling for other major factors affecting the cue. The outcome is two values per speaker: the intercept (e.g. average of VOT for voiced and voiceless stops) and slope (e.g. the difference between VOT for voiceless and voiced stops), corresponding to each speaker’s ‘overall’ cue value and the size of the voicing contrast in the cue. These two values can also be used to determine a speaker’s cue values for voiceless and voiced stops separately.

For VOT and CD, the statistical models below use the log-transformed value of the cue (log-transformed to bring model residuals closer to normality). We analyze VDC coded as a ternary variable, with levels **None**, **Some**, and **All** (0%, 1–99%, 100% percent of closure voiced), because most tokens (80.5%) had close to none or all of the closure voiced. This choice follows Davidson’s (2016; 2017) analyses of Voicing During Closure in American English stops.³

³With the proviso that Davidson (2016) uses 10–90% as the ‘some voicing’ category, rather than 1–99% as used here.

Table 2: Summary of predictors used in models for VOT, VDC, and CD.

Predictor	Type	Levels/Units	Coding
SPEECH RATE DEVIATION (VOT, CD only)	continuous	syllables/second	—
CONSONANT DURATION DEVIATION (VDC only)	continuous	log(second)	—
VOICING	factor	voiced, voiceless	sum contrasts: voiceless (1) vs. voiced (-1)
PLACE OF ARTICULATION	factor	bilabial, alveolar, velar	Helmert contrasts: POA ₁ —alveolar (1) vs. bilabial (-1); POA ₂ —velar (2) vs. bilabial/alveolar (-1)
PHRASE POSITION (VOT only)	factor	initial, medial	sum contrasts: medial (1) vs. initial (-1)
PRECEDING SEGMENT CLASS (VDC, CD only)	factor	voiceless fricative, voiced fricative, vowel	Helmert contrasts: PRECEDING SEGMENT ₁ —voiced (1) vs. voiceless (-1) fricative; PRECEDING SEGMENT ₂ —vowel (2) vs. fricative (-1)

4.2 Statistical models

We model each cue using mixed-effects models, using the `lme4` package (Bates et al., 2015) in R, including fixed-effects terms to model the effect of stop voicing (voiced vs. voiceless) and to control for other factors affecting the cue, and modeling speaker variability using random-effect terms.

4.2.1 Controls

Table 2 summarizes predictors included in the models—stop VOICING (of primary interest) and control predictors—and how each one was coded. A couple predictors merit discussion:

- (a) SPEECH RATE DEVIATION: Speech rate is defined in syllables per second within a ‘phrase’ bounded by force-aligned pauses of at least 150 ms. Speech rate was transformed to be a derivation from the speaker’s mean speech rate across the dataset, to capture effects of faster or slower speech rate on a cue’s value by a given talker within and across their utterances (see Stuart-Smith et al., 2015 for discussion for VOT).
- (b) PHRASE POSITION: Initial versus medial position in the ‘phrase’.
- (c) CONSONANT DURATION DEVIATION: Consonant duration for each stop token in the VDC dataset was approximated here as the time between the left force-aligned boundary of the stop and the end of its positive VOT annotation, measured in ms then log-transformed. Because consonant duration is effectively a very local measure of speech rate, it was transformed similarly to speech rate, by subtracting the speaker’s mean value across the VDC dataset.

Each predictor is conceptually ‘centered’, either by definition (SPEECH RATE and CONSONANT DURATION deviations) or by coding using contrasts where the intercept corresponds to the grand mean. As a result, the regression coefficient for each predictor participating in interactions in the statistical models can be interpreted as the ‘average’ effect of the predictor,

across levels of other predictors. This is relevant for the VOICING terms, which capture the difference between voiced and voiceless stops in the cue value, averaging over other variables.

4.2.2 VOT model

We fit a single linear mixed-effects model of $\log(\text{VOT})$ for all stops, voiced and voiceless. This is one of two differences from the VOT models in Stuart-Smith et al. (2015), where linear mixed-effects models were built separately for voiced and voiceless stops for the same dataset. The other difference is that the goal of the 2015 models was determining which of a range of factors that affect VOT in laboratory studies also affect VOT in spontaneous speech, while the goal of the current models was to determine each speaker’s characteristic VOT values, after controlling for major factors affecting VOT. As such, the current models only contained a subset of terms from the 2015 models.

Fixed effects were included for VOICING of the stop, as well as every term that significantly affected VOT for either voiced or voiceless stops in Stuart-Smith et al. (2015), to control for major factors affecting VOT (see Section 2.1): SPEECH RATE DEVIATION, PHRASE POSITION, and POA, as well as interactions of VOICING with these three variables, and a three-way VOICING:SPEECH RATE:POA interaction. These terms account for: VOT reduces with increasing speech rate and is lower phrase-medially than phrase-initially (each effect possibly differing by voicing); and VOT is strongly conditioned by place of articulation, with the effect modulated by speech rate and stop voicing.

For random effects, the VOT model included a by-speaker random intercept and by-speaker random slope for VOICING, as well as the correlation between them. These terms are of key interest for our goal of capturing interspeaker variation in overall VOT and the size of the voicing contrast. The model also included a by-word random intercept, to account for differences between words (beyond variables included in the model), and all possible by-speaker and by-word random slopes (Barr et al., 2013).

To avoid overparametrized models, for all statistical models (VOT, VDC₁, VDC₂, CD), correlations between random-effect terms were omitted and zero random effect terms were iteratively excluded until the fit was non-singular. The resulting random-effect structures are shown in Supplemental Materials.

4.2.3 VDC models

Conceptually, the levels of the ternary VDC variable follow an order: **None**<**Some**<**All**. We therefore use a mixed-effects ordinal regression, which models a multinomial outcome whose levels are ordered. We use a variant of this method which models two ‘continuation ratios’ (Agresti, 2002, 7.4, 12.5): the probability of one level versus the higher levels. In our case, these are two binary mixed-effects logistic regressions, one which models **None** versus **Some/All** (the probability of any closure voicing) and one which models **Some** versus **All** (the probability of full closure voicing, given that >0% of the closure is voiced). We call these the **None/Any** and **Some/All** models, or the VDC₁ and VDC₂ models. For example, the VOICING coefficient for the **None/Any** regression captures: how much higher are the log-odds of there being any Voicing During Closure, versus no Voicing During Closure, for voiced stops compared to voiceless stops? Together, these two models describe the likelihood of each VDC profile (**None/Some/All**) as a function of consonant VOICING, control predictors, and speaker/word variability.

A continuation-ratio ordinal regression is one of a family of similar methods for modeling multinomial outcomes (Gelman & Hill, 2007, Ch. 6; Agresti, 2002, Ch. 7, 12), including the more common ‘multinomial logistic regression’ used by Davidson (2016), where the probabilities of each outcome versus a fixed baseline (one level, or all levels together) are modeled. We used a continuation-ratio model because it was easier to fit and allowed us more flexibility in model specification, but our results should not differ from a multinomial logistic regression.⁴

The control predictors included in the models were PLACE OF ARTICULATION, PRECEDING SEGMENT CLASS, and CONSONANT DURATION DEVIATION—all of which were expected to affect the degree of closure voicing (see Section 2.1), and which significantly contributed to the likelihood of at least one VDC model. Other variables we considered which could plausibly affect VDC, such as phrase-level speech rate and position of the stop in the word, did not significantly improve model likelihood.

Fixed effects were included for these three predictors in both VDC models. Because there was no a priori reason to expect different effects for voiced and voiceless stops, interactions of control predictors with VOICING were included only if they significantly improved model likelihood. The **None/Any** model included VOICING:PRECEDING SEGMENT CLASS and VOICING:PLACE OF ARTICULATION terms, and the **Some/All** model included a VOICING:CONSONANT DURATION DEVIATION interaction.

For random effects, both VDC models included a by-speaker random intercept and by-speaker random slope for VOICING, as well as the correlation between them; and a by-word random intercept, all motivated identically to the VOT model. Each model included by-speaker random slopes only for fixed effects which were significant ($p < 0.05$); these terms were then pruned to give a non-singular fit, as for the VOT model.

4.2.4 CD model

We modeled CD using a linear mixed-effects model of $\log(\text{CD})$ for both voiced and voiceless stops. Fixed effects were included for VOICING and for control predictors expected to affect CD, based on previous work (see Section 2.1): PLACE OF ARTICULATION, PRECEDING SEGMENT CLASS, and SPEECH RATE DEVIATION. Fixed effects were also included for the interaction of PLACE OF ARTICULATION with VOICING (Byrd, 1993); and for the interaction of PRECEDING SEGMENT CLASS with VOICING, as this significantly improved model likelihood. We did not include any other interactions because there was no a priori reason to expect them and they did not significantly improve model likelihood.

For random effects, the model included a by-speaker random intercept and by-speaker random slope for VOICING, as well as the correlation between them; and a by-word random intercept, motivated similarly as for VOT and VDC. Each model also included all possible by-speaker and by-word random slopes, pruned to avoid a singular fit.

4.3 Results: Voicing (group level)

We first describe the size of the voicing contrast in Glasgow vernacular for each cue, using the fixed-effect results in Tables 3–5. To focus on our first research question, we do not discuss

⁴Fitting a multinomial model requires jointly fitting several component regression models for binary data. Jointly fitting models is computationally difficult and not possible using standard mixed-modeling packages in R (e.g. `lme4`). In a continuation-ratio model, the outcomes of the binary regressions are orthogonal, which means that fitting the binary regressions separately (which is easy in `lme4`) should give equivalent results to fitting them jointly (Agresti, 2002, 7.4, 12.5).

Table 3: Summary of fixed-effect coefficients in the model of $\log(\text{VOT})$. Subscripts refer to contrasts of categorical variables (Table 2).

Coefficient	Estimate	SE	df	t	$P(> t)$
Intercept	-3.616	0.032	25.7	-112.91	<0.001*
Place of articulation ₁	0.173	0.018	29.4	9.44	<0.001*
Place of articulation ₂	0.098	0.011	37.5	9.28	<0.001*
Voicing	0.522	0.026	27.9	20.20	<0.001*
Speech rate (deviation)	-0.018	0.006	19.8	-3.03	0.007*
Phrase position	-0.044	0.010	23.7	-4.24	<0.001*
POA ₁ :Voicing	-0.021	0.014	25.8	-1.49	0.149
POA ₂ :Voicing	-0.037	0.009	38.4	-4.08	<0.001*
Voicing:Speech rate	-0.005	0.008	23.9	-0.65	0.519
Voicing:Phrase position	0.018	0.013	21.0	1.39	0.18
POA ₁ :Speech rate	-0.004	0.006	26.0	-0.59	0.557
POA ₂ :Speech rate	-0.008	0.004	24.7	-2.12	0.044*
POA ₁ :Voicing:Speech rate	-0.003	0.007	24.4	-0.48	0.637
POA ₂ :Voicing:Speech rate	0.003	0.004	24.8	0.81	0.425

*indicates $p < 0.05$

the results for control predictors (i.e. predictors besides VOICING) or how these predictors modulate the VOICING effect; full discussion is given in Supplemental Materials. For each cue, the control predictors largely affect the cue in ways expected from previous work on (mostly) read speech in standard varieties of English, giving confidence in the quality of our measures for spontaneous speech in a vernacular variety.

Fixed-effect coefficients are shown with significances calculated using the Satterthwaite approximation using `lmerTest` (Kuznetsova et al., 2017) for VOT and CD (Tables 3, 5), and calculated with a Wald test for VDC models (Table 4).

Voiced stops have significantly lower (log-transformed) VOT than voiceless stops (VOICING: $\hat{\beta}=0.52$, $t=20.2$, $p < 0.001$), corresponding to a voiced/voiceless difference of 29 ms (mean VOT values in the dataset: voiced=18 ms; voiceless=50 ms), averaging over other variables.⁵

For VDC, voiced stops are intuitively ‘more voiced’ than voiceless stops: some amount of voicing during the closure is more likely for voiced stops (**None/Any** model VOICING: $\hat{\beta}=-0.51$, $z=-4.04$, $p < 0.001$), as is full closure voicing compared to partial voicing (**Some/All** model VOICING: $\hat{\beta}=-0.24$, $z=-1.9$, $p=0.06$), though the latter result is above the significance threshold. Averaging over other variables, no/partial/full closure voicing is predicted to occur for 78.3%/8.7%/13% of voiceless stops (proportions in dataset: 56.9%, 20%, 23%), and 56.8%/12.6%/30.6% of voiced stops (proportions in dataset: 27.6%, 19%, 53.3%).⁶

⁵Note that the VOT and CD models only predict log-transformed values, and the VDC models only predict the relative odds of each VDC outcome (**None/Some/All**). To predict differences in ms (VOT, CD) or probabilities (VDC), it is necessary to first fix values of each control predictor. We calculate predictions at each model’s intercept throughout Analysis 1.

⁶About 80% of tokens are post-vocalic (where closure voicing is more likely), while model predictions average over all preceding contexts (vowel, voiced fricative, voiceless fricative), leading to the discrepancy between empirical proportions and model predictions.

Table 4: Summary of fixed-effect coefficients in models of Voicing During Closure. Coefficient subscripts refer to contrasts of categorical variables (Table 2).

VDC₁ model: **None** (0) vs. **Some/All** (1)

Coefficient	Estimate	SE	z	$P(> z)$
Intercept	-0.78	0.14	-5.48	<0.001*
Voicing	-0.51	0.13	-4.04	<0.001*
Preceding segment class ₁	0.56	0.19	3.01	0.003*
Preceding segment class ₂	0.81	0.08	10.41	<0.001*
Place of articulation ₁	0.16	0.07	2.36	0.018*
Place of articulation ₂	-0.05	0.04	-1.19	0.232
Consonant duration (deviation)	0.42	0.21	1.98	0.048*
Voicing:Preceding segment ₁	-0.43	0.15	-2.98	0.003*
Voicing:Preceding segment ₂	-0.26	0.05	-4.97	<0.001*
Voicing:POA ₁	0.02	0.07	0.32	0.746
Voicing:POA ₂	0.12	0.04	3.18	0.001*

VDC₂ model: **Some** (0) vs. **All** (1)

Coefficient	Estimate	SE	z	$P(> z)$
Intercept	0.65	0.26	2.50	0.013*
Voicing	-0.24	0.13	-1.90	0.057
Preceding segment class ₁	0.00	0.29	0.01	0.994
Preceding segment class ₂	0.06	0.10	0.59	0.558
Place of articulation ₁	0.04	0.08	0.52	0.602
Place of articulation ₂	-0.02	0.05	-0.37	0.715
Consonant duration (deviation)	-1.70	0.32	-5.28	<0.001*
Voicing:Consonant duration	0.84	0.36	2.37	0.018*

*indicates $p < 0.05$

Table 5: Summary of fixed-effect coefficients in the model of $\log(\text{Closure Duration})$. Subscripts refer to contrasts of categorical variables (Table 2).

Coefficient	Estimate	SE	df	t	$P(> t)$
Intercept	-2.986	0.022	27.4	-134.07	<0.001*
Voicing	0.057	0.011	60.6	5.10	<0.001*
Place of articulation ₁	-0.044	0.019	20.7	-2.34	0.029*
Place of articulation ₂	-0.032	0.007	23.3	-4.69	<0.001*
Preceding segment class ₁	-0.003	0.017	14.8	-0.19	0.853
Preceding segment class ₂	0.020	0.009	21.7	2.32	0.03*
Speech rate (deviation)	-0.015	0.005	17.6	-2.82	0.012*
Voicing:POA ₁	-0.037	0.016	23.2	-2.35	0.028*
Voicing:POA ₂	0.003	0.005	26.3	0.52	0.607
Voicing:Preceding segment ₁	-0.031	0.013	1261.9	-2.31	0.021*
Voicing:Preceding segment ₂	0.002	0.005	994.6	0.50	0.616

*indicates $p < 0.05$

Table 6: Subset of random-effect terms from the models of VOT, VDC, and CD: by-speaker random intercept and random slope for VOICING (reported as a standard deviation: σ), and correlation between them.

	Intercept σ	VOICING σ slope	Correlation
VOT	0.143	0.11	-0.229
VDC ₁	0.405	0.237	0.035
VDC ₂	0.752	0.357	0.324
CD	0.09	0.02	0.311

For CD, voiceless stops have significantly higher (log-transformed) Closure Duration than voiced stops, averaging across other variables (VOICING: $\hat{\beta}=0.06$, $t=5.1$, $p < 0.001$), corresponding to a voiced/voiceless difference of 5 ms (mean CD values in the dataset: voiced=49 ms; voiceless=53 ms).

Thus, aggregated across speakers, places of articulation, etc., there is a clear difference between voiced and voiceless stop categories in the expected direction, for each cue.

4.4 Results: Individual speaker variability

To address our first research question, we unpack the relevant aspects of these random effects—intercept and by-speaker random slope of VOICING terms for each cue (shown in Table 6)—in three ways. The full random-effect tables are given in Supplemental Materials.

We first ask whether speakers significantly differ in overall use of each cue and in contrast size. For VOT, speakers significantly differ in both ways, as assessed by a likelihood ratio test comparing models with and without each term: in overall $\log(\text{VOT})$ ($\chi^2(2) = 119.4$, $p < 0.001$), and in the contrast ($\chi^2(2) = 199.7$, $p < 0.001$).⁷ For Voicing During Closure, speakers differ significantly in the overall degree of VDC in both models (**None/Any** voicing: $\chi^2(2) = 9.5$, $p = 0.009$; **Some/All** voicing: $\chi^2(2) = 66.9$, $p < 0.001$), reflecting large differences among speakers in the degree of closure voicing, across all stops (voiced and voiceless). In terms of the contrast, speakers differ significantly for the **None/Any** voicing model ($\chi^2(2) = 9.4$, $p = 0.009$), while for the **Some/All** voicing model they show a non-significant trend ($\chi^2(2) = 5.8$, $p = 0.055$). For Closure Duration, speakers significantly differ in overall $\log(\text{CD})$ ($\chi^2(2) = 199.7$, $p < 0.001$), but do not significantly differ in the contrast ($\chi^2(2) = 2.3$, $p = 0.315$).

Second, it is useful to consider the size of predicted interspeaker differences, which can be calculated using the by-speaker intercept and by-speaker random VOICING slope terms from the statistical model for each cue; the intercept plus or minus twice the random slope value describes the range of values for 95% of speakers.

For VOT, most speakers (95%) have an overall value between 20 and 36 ms, compared to the population mean of 27 ms, and a voiced/voiceless VOT contrast of 16–44 ms. For VDC, for simplicity we only quantify the degree of interspeaker variability in whether any VDC is present (**None/Any** model). Most speakers show some degree of Voicing During Closure between 17% and 50.7% of the time, compared to the population mean of 31.4%. For the contrast, speakers vary roughly between 0.01 and 0.4 difference in probability of **Any** voicing. For CD, most speakers have an overall value between 42 ms and 60 ms, compared to the population mean of 50 ms. Speakers do not differ significantly in CD contrast size.

Third, we can extract estimates of the predicted cue values for voiced and voiceless stops for each speaker (‘BLUPs’: Pinheiro & Bates, 2000, Sec. 2.2; Gelman & Hill, 2007), controlling for other variables, to get a sense of what interspeaker differences are predicted at the level of individuals. For VOT (Figure 2: top left), we see that despite substantial interspeaker variability in the overall value, each speaker makes a clear **CONTRAST**, with VOT higher for voiceless than for voiced stops.

For VDC, Figure 3 shows the predicted probability of each voicing class (**None/Some/All**), for each speaker, for voiced and voiceless stops. There are large differences in how much speakers voice during closure overall, as reflected in the different heights of the **None** and **All** lines in particular. Despite this variability, there is a clear contrast for each speaker: there

⁷For example, the test for ‘overall $\log(\text{VOT})$ ’ excludes the by-speaker random intercept and its correlation with the by-speaker random VOICING slope.

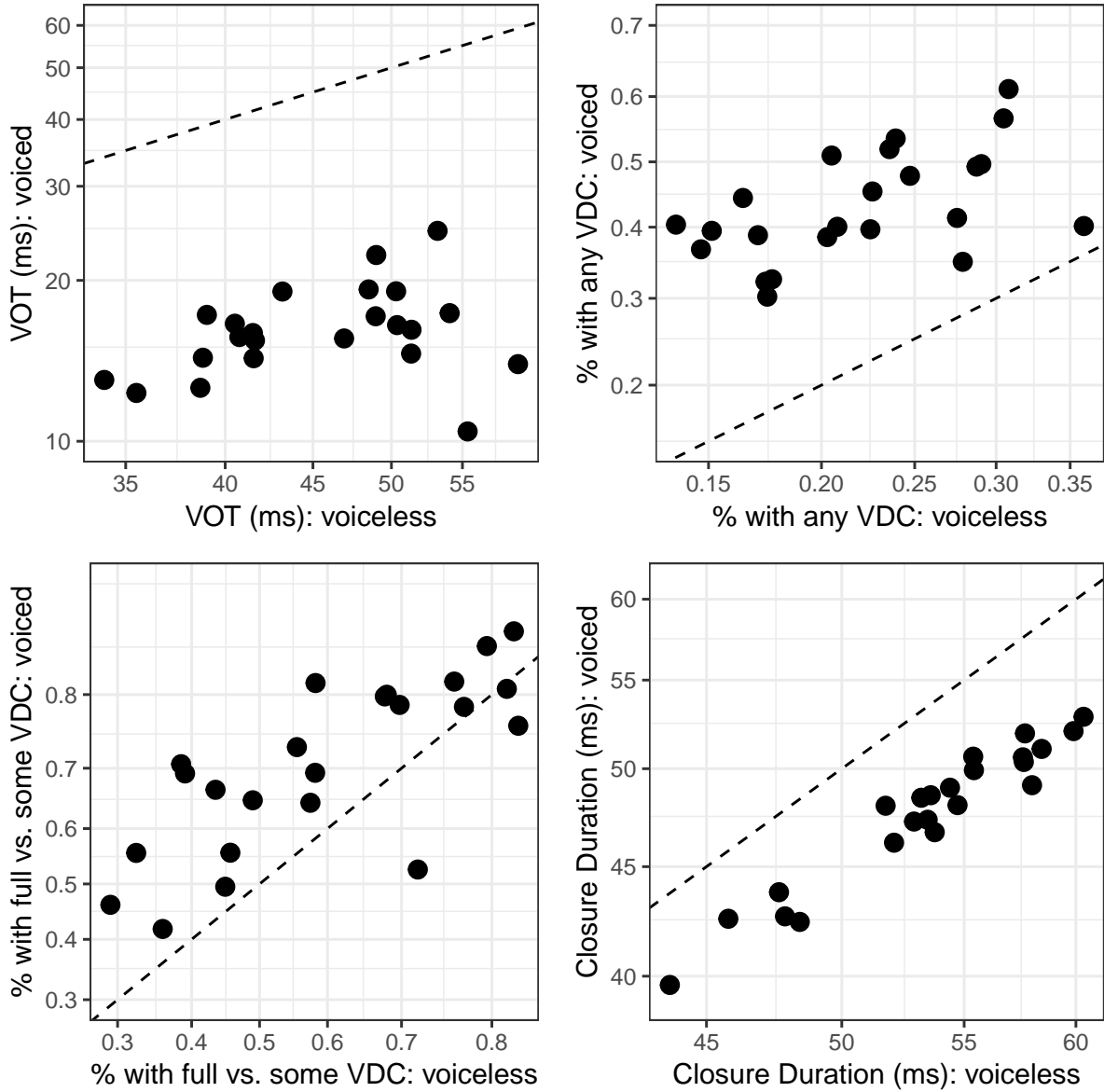


Figure 2: Model-predicted cue values for voiceless and voiced stops for each speaker (one point per speaker): VOT (top-left), Voicing During Closure (top-right: $P(\text{Some})+P(\text{All})$; bottom-left: $P(\text{All})/(P(\text{Some})+P(\text{All}))$), and Closure Duration (bottom-right). Dotted line is $y=x$, i.e. where the value for voiced stops equals that for voiceless stops. VOT and CD on log scale, VDC probabilities on log-odds scale.

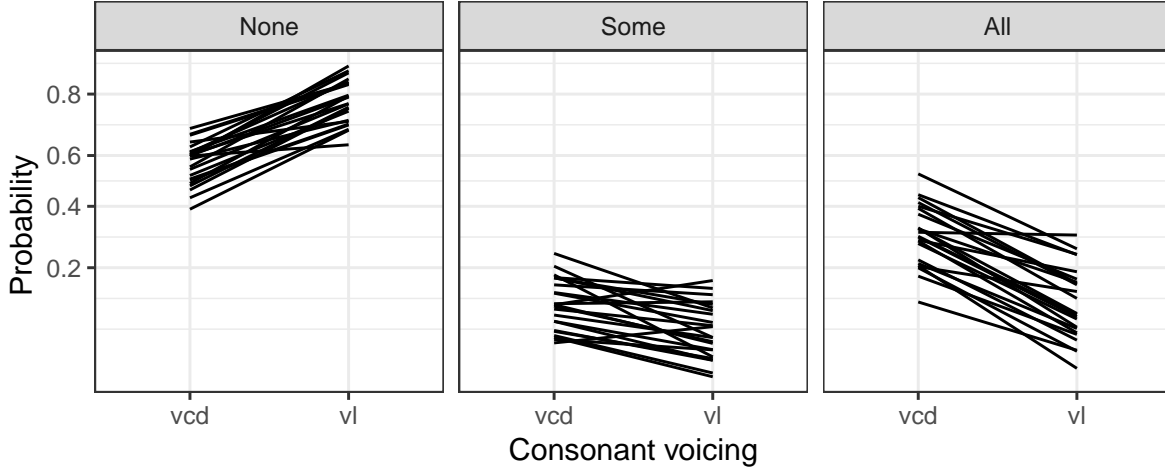


Figure 3: Model-predicted probabilities of each VDC class (*None*, *Some*, *All*) for voiced and voiceless stops, for each speaker (one line per speaker). On log-odds scale.

is less *None* and more more *All* voicing for voiced stops than for voiceless stops (positive and negative slope of *None*, *All* lines). In other words, each speaker broadly shows more voicing during the closure for voiced stops than for voiceless stops. To visualize the significant variability in how speakers use VDC, Figure 2 (top-right, bottom-left) shows each speaker’s predicted probability of any voicing and of full vs. partial voicing, for voiced and voiceless stops. For both measures, speakers generally lie along a continuum from ‘less Voicing During Closure’ to ‘more Voicing During Closure’, across voiced and voiceless stops. Figure 4 shows that the same speakers who use any Voicing During Closure more (for the average of voiced and voiceless stops) also tend to use full voicing more relative to partial voicing. In other words, speaker variability for VDC largely lies along a continuum, of those who show ‘less voiced’ to those who show ‘more voiced’ closures, across all stops. To a lesser degree speakers also vary along an orthogonal dimension: how large the voiced/voiceless contrast is in showing **Any Voicing During Closure** (VDC_1); this is the distance from the dotted line in Figure 2 (top-right).

For Closure Duration (Figure 2: bottom-right), we see that all speakers maintain the contrast, with higher CD for voiceless than for voiced stops. Speakers basically lie along a line of increasingly long Closure Duration. The fact that each speaker signals the stop voicing using CD is striking given the very small magnitude of the contrast: voiced and voiceless stops differ in Closure Duration by only 5 ms on average.

4.5 Analysis 1: Summary

Our first research question was whether speakers differ in their use of each phonetic cue to realize the stop voicing contrast in this spontaneous speech dataset. After controlling for key linguistic and prosodic factors for each cue (VOT, VDC, CD), we find that speakers do show substantial differences in their ‘overall’ use of each cue—reflecting structured variability in how speakers realize the voiced and voiceless categories for each cue. Speakers also differ to a smaller extent in the size of the contrast they make with each cue (for VOT and VDC only). At the same time, even in this least-controlled speech style, every speaker maintains

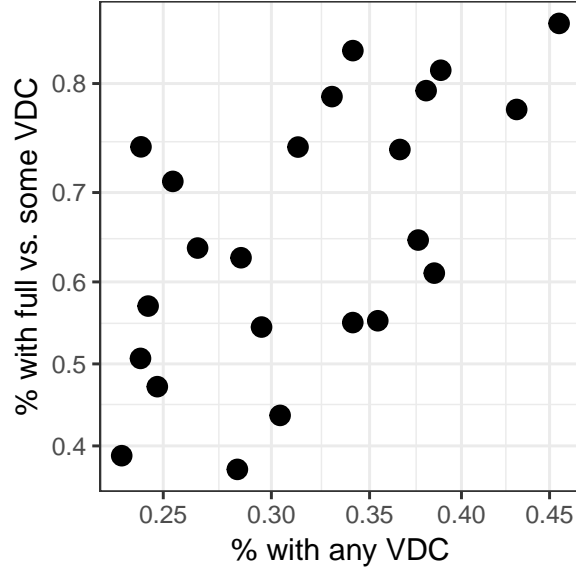


Figure 4: Model-predicted probability of any Voicing During Closure ($=P(\text{Some})+P(\text{All})$) and of full vs. partial voicing ($=P(\text{All})/(P(\text{Some})+P(\text{All}))$), for each speaker (one point per speaker) for all stops (average of voiced and voiceless). On log-odds scales.

the stop voicing contrast in the expected direction for each cue.

5 Analysis 2: Individual speaker variation across phonetic cues

We have shown that individual speakers differ in their use of each cue to the stop voicing contrast, and also in how they realize this contrast using VOT and the presence of **Any** Voicing During Closure. We now ask: is the use of individual speakers' cues correlated in the realization of the stop voicing contrast?

The models fitted above resulted in a description of how individuals realize the contrast, as 'intercept' (average of voiced and voiceless) and 'slope' (voiceless minus voiced) values, for each cue. Individual speaker variability can be described in terms of three sets of coefficients:⁸

1. VOT: intercept, slope (voiceless-voiced)
2. VDC: intercept, slope (voiced-voiceless) for **None** vs. **Any** VDC; intercept, slope for **Some** vs. **All** VDC
3. CD: intercept, slope (voiceless-voiced)

We now assess for each PAIR of cues (VOT, VDC, CD), whether interspeaker variability in the two cues is correlated, and how. One way of doing this would be to examine every possible correlation (e.g. VOT intercept with VDC₁ slope, etc.), but this method presumes that intercept and slope are the right variables to consider. In the absence of any previous

⁸Note that 'slope' is defined differently for VDC than for VOT/CD, so that its expected value is positive for all speakers (higher VDC for voiced stops; higher VOT/CD for voiceless stops).

Table 7: Summary of CCA results for each pair of cues: correlation between the components for each cue, and significance of the correlation as assessed by a permutation test (using test statistic Wilks’ lambda).

	VOT/VDC	VOT/CD	VDC/CD
Correlation	-0.8	0.54	-0.67
P value	0.015*	0.139	0.115

*indicates $p < 0.05$

Table 8: Weights of the CCA components for the VOT/VDC cue pair.

VOT component		VDC component			
Intercept	Slope	VDC ₁ Intercept	VDC ₁ Slope	VDC ₂ Intercept	VDC ₂ Slope
0.99	-0.12	-0.3	0.94	0.17	0.05

work which considers the relationships between the cues, we do not know this. While we have good reason to suspect some kind of interspeaker correlation could exist based on previous work (see Section 2.3), our search for interspeaker correlations is fundamentally exploratory, and should be as flexible as possible. We therefore ask for each pair of sets (1)–(3): what is the linear combination of intercept(s) and slope(s), for each cue, that is best correlated with some combination of the intercept(s) and slope(s) for the other cue? This question is answered by CANONICAL CORRELATION ANALYSIS (CCA; González et al., 2008). CCA gives, for two sets of variables measured on the same set of observations (here, individual speakers), a linear combination for each set of variables (‘component’) that show the maximum correlation. Since some such correlation will likely be observed by chance, a permutation test can be used to assess the significance of the association between the two sets of variables.

We carried out a CCA analysis for each pair of sets (1)–(3): VOT and VDC (are speakers’ use of VOT and VDC correlated?), and so on. Table 7 shows, for each pair, the correlation between the two best-correlated components, and its significance. There is a strong and significant correlation between speakers’ use of VOT and of VDC, the two cues for which speakers show the largest contrast ($r = 0.8$, $p = 0.015$). Speakers’ use of Closure Duration and VOT, as well as Closure Duration and VDC, are not significantly correlated.

To understand what aspects of cue use are correlated across individual speakers, we can examine the structure of the primary component (the weight of ‘intercept’ and ‘slope’ terms) for VOT and VDC, for the VOT/VDC correlation (Table 8). The VOT component is largely interpretable as ‘overall VOT’, with a slight negative weight for ‘VOT contrast’. The VDC component is largely interpretable as ‘contrast in VDC₁’, with some negative weight for ‘overall VDC₁’. Thus, the negative correlation between these two components can be interpreted primarily as: speakers who have a longer overall VOT value also use the presence of Voicing During Closure (none vs. any) LESS to signal the voicing contrast (smaller difference between voiced and voiceless stops). This more intuitive relationship between VOT and VDC₁, shown in Figure 5, is the interpretation that we assume going forward.

This negative relationship is itself strong ($r = -0.74$, $p < 0.001$). The point of carrying

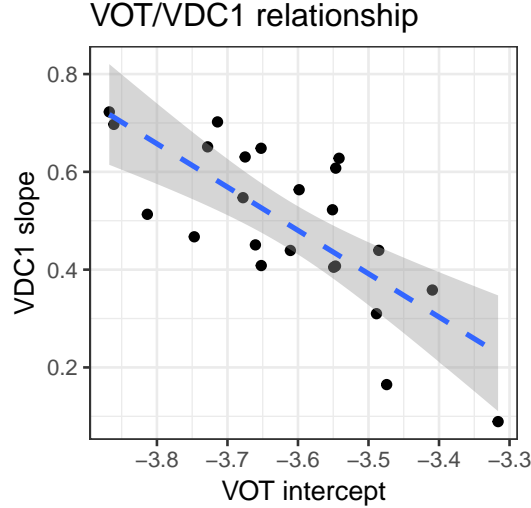


Figure 5: Correlation between speakers’ VOT Intercept and VDC_1 slope values ($r = -0.74$), corresponding approximately to the relationship between the two cues found via CCA. Each point corresponds to one speaker. Dotted line and shading is line of best fit with 95% CIs.

out a CCA analysis, rather than just reporting this relationship, is to guard against finding such a relationship by chance, given all the possible ways speakers’ VOT and VDC values could be correlated. For the VOT/CD and VDC/CD relationships, it is possible to find ways that speakers use the two cues which are still relatively strongly correlated (the top row of Table 7) in our sample. Nonetheless, the CCA analysis says there is not enough evidence in our data to conclude that these relationships exist. CCA is a conservative statistical method; if anything we are underestimating the degree of correlation between speakers’ use of different cues.

5.1 Analysis 2: Summary

Analysis 2 answers our second research question: beyond speaker variation in how single cues are used independently (Section 4), speakers display structured variability in how they modulate cues to signal the voicing contrast. Namely, speakers who produce more aspirated stops (greater positive VOT) use Voicing During Closure less to signal the contrast. Our final analysis considers whether including the additional factor of time illuminates a further diachronic layer of structure for these speakers.

6 Analysis 3: Across cues, across speakers, over time

Analyses 1 and 2 show substantial structured interspeaker variability in how the Glaswegian voicing contrast is realized in spontaneous speech. Speakers vary greatly in the use of all three cues, and speakers’ use of two of these cues is correlated. What could explain why individuals vary in overall values and contrast sizes for a given phonetic cue? Working under the assumption of structured heterogeneity binding synchronic and diachronic lan-

guage description together (Weinreich et al., 1968), several possibilities present themselves, including contrast maintenance, differences in speaking style, or social factors (e.g. gender, identity construction); here we examine the possibility of CHANGE OVER TIME (Bang, 2017). In Analysis 3, we ask: is the use of individual speakers’ cues for the stop voicing contrast structured according to decade of recording, and so consistent with change over time?

6.1 Analysis 3: Method

We address this question through quantitative (Sections 6.2, 6.3, 6.4) and qualitative (Section 6.5) analyses. The quantitative analyses use statistical models which consider speakers in terms of groups by decade of birth. Recall that speakers in our sample fall into six groups, corresponding to old, middle-aged, and young speakers from two recording decades (1970s, 2000s); Table 1. We define the factor DECADE OF BIRTH to track which group a speaker belongs to.

For each model of the phonetic cues described in Analysis 1, we fit an identical model, but with fixed-effect terms added for DECADE OF BIRTH and its interaction with VOICING.⁹ This tests whether there has been change in the overall cue value, and its use in the voicing contrast, over time. The updated model now estimates the cue value for voiced and voiceless stops for speakers with each DECADE OF BIRTH, after accounting for controls and variability between speakers (BEYOND decade of birth). The updated model is then used to estimate the degree of REAL-TIME change in each age cohort, in overall cue use and in contrast size, giving a total of six estimates:

1. Change in ‘average of voiced and voiceless’ for Old speakers from 1970s to 2000s.
2. Change in ‘voiceless minus voiced’ for Old speakers from 1970s to 2000s
3. Change in ‘average of voiced and voiceless’ for Middle-aged speakers from 1970s to 2000s
4. Change in ‘voiceless minus voiced’ for Middle-aged speakers from 1970s to 2000s
5. Change in ‘average of voiced and voiceless’ for Young speakers from 1970s to 2000s
6. Change in ‘voiceless minus voiced’ for Young speakers from 1970s to 2000s

For each model, we estimate these six values using the `emmeans` package (Lenth, 2018). To assess whether the estimates are significantly different from zero, corresponding *p*-values are calculated also using `emmeans` (using Satterthwaite approximation/Wald tests as for the mixed-effects models).

6.2 VOT: By group over time

Stuart-Smith et al. (2015) conducted a real- and apparent-time analysis of this VOT dataset. This analysis did not consider voiced and voiceless stops TOGETHER, as we do here, to assess change in overall VOT value and contrast size. For simplicity of presentation and comparability with the VDC and CD analyses, the current analysis does not account for interactions with place of articulation.

⁹Note that the models including DECADE OF BIRTH could not themselves be used for Analysis 1, because inclusion of this term means the by-speaker random effects no longer have the interpretation needed for Analysis 1 (each speaker’s characteristic value for the cue’s intercept/slope).

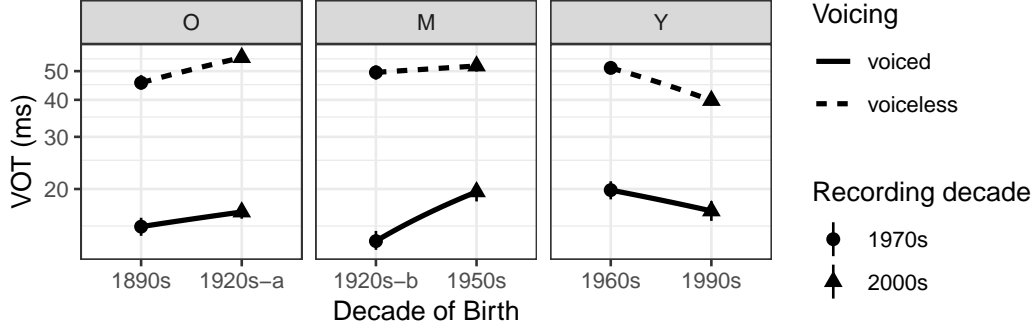


Figure 6: Mean VOT values (on log scale) for each Decade of Birth, for voiced and voiceless stops. Each point is the mean across average VOT values for each speaker/word pair (bars: 95% confidence intervals, which are not visible). Real-time comparisons are for the same age group (O=Old, M=Middle-aged, Y=Young) between 1970s and 2000s recording decades. On log scale.

Table 9: Estimated real-time differences in VOT for the overall value (average of voiced and voiceless) and the contrast (voiceless – voiced difference). Each row shows the estimated difference in $\log(\text{VOT})$, with associated standard error, t statistic, and significance.

Age group	Type	Estimated difference	SE	df	t	$P(> t)$
Old	Overall	0.215	0.089	16.5	2.428	0.027*
	Contrast	0.092	0.085	16.3	1.081	0.296
Middle-aged	Overall	0.172	0.083	17.8	2.066	0.054
	Contrast	-0.126	0.081	18.1	-1.559	0.136
Young	Overall	-0.226	0.082	16.9	-2.745	0.014*
	Contrast	0.002	0.079	16.8	0.024	0.981

*indicates $p < 0.05$

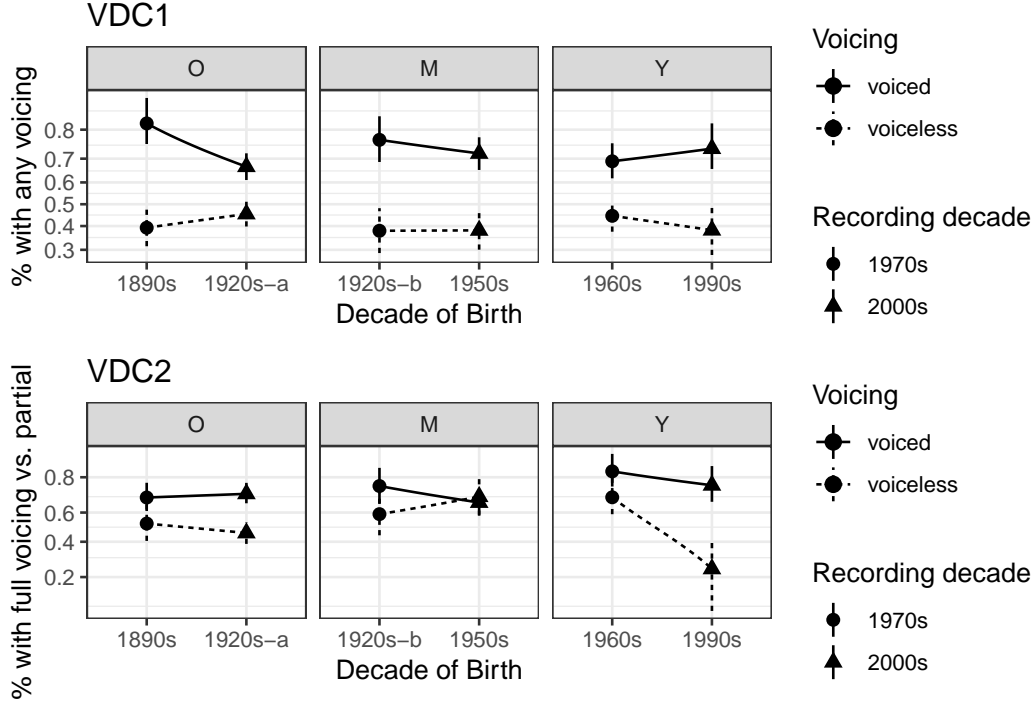


Figure 7: Mean proportion of each kind of Voicing During Closure for each Decade of Birth, for voiced and voiceless stops. Top: any voicing ($=P(\text{Some})+P(\text{All})$). Bottom: all voicing versus partial voicing ($=P(\text{All})/(P(\text{Some})+P(\text{All}))$). Each point is the mean across average proportions for each speaker/word pair (bars: 95% confidence intervals). Real-time comparisons are as in Figure 6. On log-odds scale.

Figure 6 shows empirical trends in VOT by Decade of Birth, for voiced and voiceless stops. As we consider possible evidence for real-time change in each cue, it is useful to refer to the empirical trend plot to understand the corresponding model's predictions. Real-time comparisons for a given age group correspond to comparing the left and right sides of an age panel.

Table 9 shows the estimated difference in $\log(\text{VOT})$ for each real-time comparison. Overall VOT significantly increased for Old speakers (est. diff.=0.21, $p = 0.027$) and significantly decreased for Young speakers (est. diff.=−0.23, $p = 0.014$) over time, while there is a non-significant tendency for VOT to increase for Middle-aged speakers (est. diff.=0.17, $p = 0.054$), all reflecting the trends observed in the empirical data. The VOT contrast does not significantly change over time for any age group ($p > 0.136$).

Thus, VOT increases over real time for Old speakers (and possibly Middle-aged speakers), decreases over time for Young speakers, and shows no significant change in contrast size. Recall that speakers differed significantly in both overall VOT and in contrast size (Analysis 1). Thus, how VOT varies over time is a subspace of how it varies among speakers in the population, along the axis of ‘how aspirated’ stops are.

Table 10: Estimated real-time differences in Voicing During Closure for the overall value (average of voiced and voiceless) and the contrast (voiceless – voiced difference). Each row shows the estimated difference in log-odds for VDC₁ or VDC₂ model, with associated standard error, z statistic, and significance.

VDC ₁ model: None vs. Some/A11						
Age group	Type	Estimated difference	SE	z	$P(> z)$	
Old	Overall	-0.584	0.381	-1.532	0.126	
	Contrast	0.516	0.157	3.281	0.001*	
Middle-aged	Overall	0.139	0.380	0.366	0.714	
	Contrast	0.434	0.201	2.159	0.031*	
Young	Overall	-0.352	0.372	-0.949	0.343	
	Contrast	-0.298	0.210	-1.424	0.154	

VDC ₂ model: Some vs. A11						
Age group	Type	Estimated difference	SE	z	$P(> z)$	
Old	Overall	-0.541	0.599	-0.903	0.367	
	Contrast	0.082	0.220	0.372	0.71	
Middle-aged	Overall	0.171	0.579	0.296	0.767	
	Contrast	0.483	0.323	1.494	0.135	
Young	Overall	-1.085	0.563	-1.926	0.054	
	Contrast	-0.801	0.291	-2.748	0.006*	

*indicates $p < 0.05$

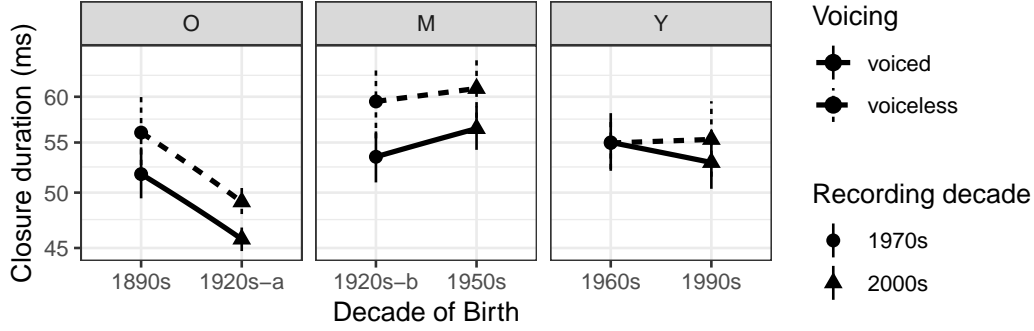


Figure 8: Mean Closure Duration values (on log scale) for each Decade of Birth, for voiced and voiceless stops. Each point is the mean across average CD values for each speaker/word pair (bars: 95% confidence intervals). Real-time comparisons are as in Figure 6. On log scale.

6.3 VDC: By group over time

Figure 7 shows empirical trends in the degree of Voicing During Closure for voiced and voiceless stops, by Decade of Birth, presented as the two proportions modeled in the VDC_1 and VDC_2 models: None vs. Any voicing, and Some vs. All voicing.

Table 10 presents the estimated differences in log-odds for these two proportions for real-time comparison, of overall VDC and the contrast. The overall probability of any voicing (VDC_1) did not significantly change for any age group ($p > 0.13$). The size of the voiceless/voiced difference decreased for Old speakers ($p = 0.001$) and Middle-aged speakers ($p = 0.03$), in line with the empirical trends in Figure 7 (top). For the probability of full versus partial voicing (VDC_2), only Young speakers show significant change: the voiceless/voiced difference strongly increases ($\beta = -0.8$, $p = 0.006$) (corresponding to a negative change, because voiceless < voiced), and the overall probability of full voicing shows a non-significant tendency to decrease (est. diff. = -1.08, $p = 0.054$). Both changes are essentially due to a large decrease in the probability of full voicing for voiceless stops (Figure 7 lower-right).

In sum, Voicing During Closure is used less over real-time for the contrast by Old and Middle-aged speakers, and used more use for the contrast by Young speakers. The overall use of VDC, across voiced and voiceless stops, shows little change. Recall that speakers differed significantly in ‘how voiced’ stop closures were overall, and showed a small difference in the contrast size for VDC_1 (and possibly VDC_2). Thus, how VDC varies over time is a subspace of how it varies among speakers in the population.

6.4 CD: By group over time

Figure 8 shows empirical trends in Closure Duration (log-transformed) for voiced and voiceless stops, by Decade of Birth. Table 11 presents the estimated differences in overall log(CD) and the contrast for each real-time comparison, with significances calculated as for VOT. The overall Closure Duration significantly decreases for Old speakers ($\beta = -0.27$, $p < 0.001$), as is reflected in the empirical data (Figure 8 left). Neither changes in overall CD for Middle-aged and Young speakers, nor change in the contrast for any age group, reach significance. Thus, there is little change over time in CD (Old speakers only); what change there is occurs along the same direction as variation across speakers (Analysis 1), where speakers differed significantly in overall CD value but not in contrast size.

Table 11: Estimated real-time differences in Closure Duration (on log scale), as in Table 9.

Age group	Type	Estimated difference	SE	df	t	$P(> t)$
Old	Overall	-0.267	0.050	14.8	-5.295	<0.001*
	Contrast	-0.005	0.026	9.9	-0.203	0.843
Middle-aged	Overall	-0.004	0.047	18.5	-0.081	0.936
	Contrast	0.020	0.030	24.9	0.684	0.5
Young	Overall	-0.079	0.047	17.6	-1.680	0.111
	Contrast	-0.029	0.032	27.8	-0.925	0.363

*indicates $p < 0.05$

6.5 By individuals, across cues, over time

Analysis 3 shows some evidence for real-time change for two cues, primarily VOT and Voicing During Closure (mainly **None** vs. **Any**). The question remains as to what is happening for individual speakers over time. Figure 9 shows a figure similar to Figure 5, in a series of panels, plotting each speaker’s estimated overall VOT value (in ms) versus how much more likely any voicing during the closure is (odds ratio) for voiced than for voiceless stops—which is a more interpretable way to think about the contrast in VDC_1 . Thus, a speaker to the top left of a panel has less aspirated stops and uses Voicing During Closure more to signal the contrast, and vice versa. Like Purnell et al. (2005), we provide a qualitative, visual display of diachronic and synchronic information for individual speakers together, which also shows the shift in the ‘trading relation’ between the cues over time.

Panel 1 shows the earliest-born group (circles), Old speakers born in the 1890s, recorded in the 1970s. These three speakers show the most use of Voicing During Closure (largest voiced/voiceless difference), and the least aspiration, although with interspeaker differences in how conservative they are. Panel 2 adds the Old speakers born in the 1920s, recorded in the 2000s (triangles). There is a clear real-time shift both in the reduction of use of Voicing During Closure and in an increase in aspiration, again with interspeaker differences. Panel 3 adds the Middle-aged speakers born in the 1920s, recorded in the 1970s (filled squares). This apparent-time shift (Old→Middle-aged, recorded in 1970s) is less noticeable than the real-time one: there is one very conservative speaker, who shows high use of Voicing During Closure and very little aspiration, but the other speakers use Voicing During Closure relatively less and show more aspiration. Panel 4 adds Middle-aged speakers born in the 1950s, recorded in the 2000s (crosses), thus showing the real-time comparison for Middle-aged speakers. Two of these speakers use less Voicing During Closure and more aspiration, one of them the most so far, whilst two are more conservative, showing less aspiration or more use of Voicing During Closure, though not as much as the conservative earliest-born speakers. Panel 5 adds the Young speakers born in the 1960s and recorded in the 1970s (squares+crosses). Again, there is a ‘spread’ pattern for these four speakers: one is advanced, showing the least use of Voicing During Closure and the most aspirated stops so far. The other three speakers show some use of Voicing During Closure, again not as much as the conservative earliest-born speakers, and different degrees of aspiration. Panel 6 adds the most recently-born group, Young speakers born in the 1990s recorded in the 2000s (stars); they revert to much less aspirated stops and more usage of closure voicing, similarly to the conservative earliest-born speakers.

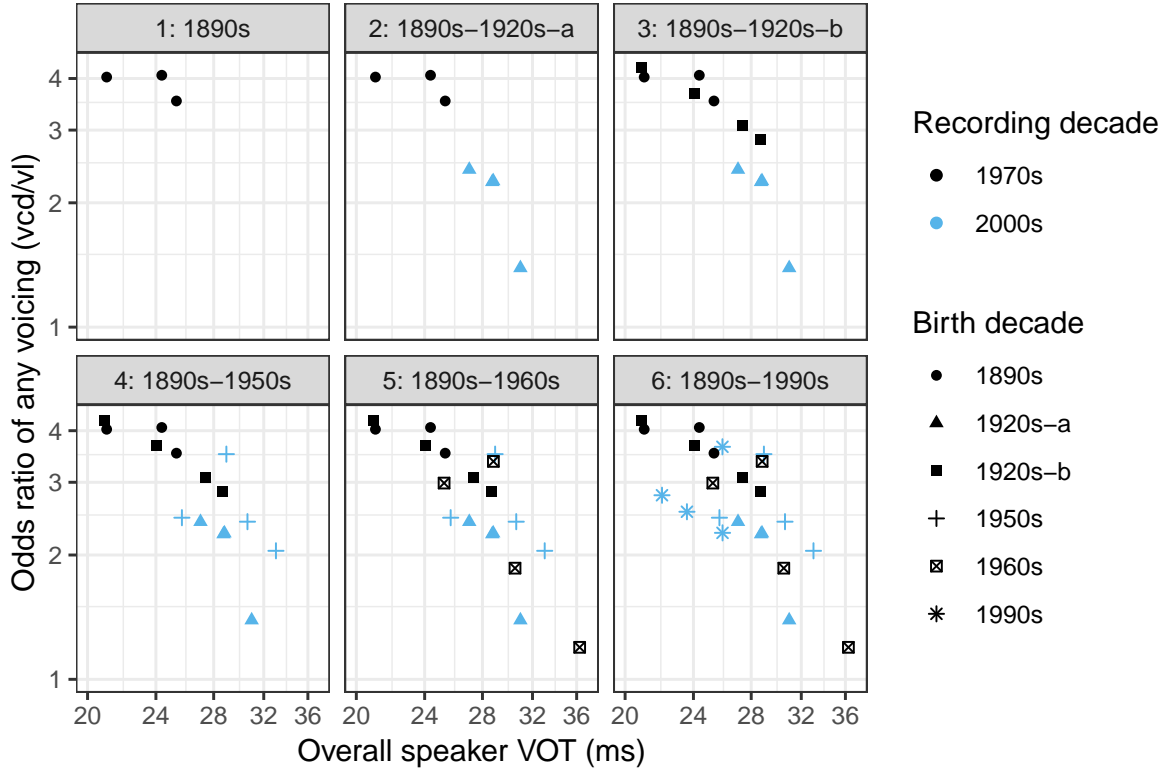


Figure 9: Model-predicted Overall VOT value and odds ratio of any Voicing During Closure for voiced versus voiceless stops (one point per speaker). Each panel shows speakers born over a progressively greater timespan (range of birth decades). (See Table 1.) Both axes on log scale.

This reveals the structured heterogeneity for this contrast, specifically two layers of description together: (1) without diachronic information about speakers, we find that speakers lie along a continuum roughly from ‘more aspirated stops, less use of voicing in contrast’ to ‘less aspirated stops, more use of voicing in contrast’. This reflects one ‘axis’ of synchronic variability for this community, in addition to other axes shown in Analysis 1 (e.g. speaker variability in how voiced all stops are). (2) Diachronic information about our speakers shows additionally that what change is taking place is also located along the aspiration/voicing axis. In other words, the two patterns are integrally connected such that the trading relation inherent in the structured variability provides the basis for diachronic change.

7 Discussion

This study considers for the first time how individual speakers control single and multiple cues to signal the stop voicing contrast over time in spontaneous speech, specifically VOT alongside two other cues to stop voicing, Voicing During Closure and Closure Duration. We structure our discussion around our research questions: the evidence for structured variability in speakers’ use of single (Section 7.1) and multiple (Section 7.2) phonetic cues, and additional structure with respect to time (Section 7.3).

7.1 Structured variability in single phonetic cues for Scottish stop voicing

Substantial interspeaker differences have been found in several cues for American English word-initial stops in different kinds of read speech, specifically VOT (Allen et al., 2003; Chodroff & Wilson, 2017, 2018), following vowel duration, and onset f_0 (Clayards, 2018). These interspeaker differences are further structured, such that each speaker maintains stop contrasts for place of articulation and voicing (Chodroff & Wilson, 2017, 2018; Theodore et al., 2009). Until now, structured variability for these cues has not been examined for spontaneous English, and Voicing During Closure and Closure Duration have not been considered at all.

We find significant differences between individual speakers’ use of VOT, presence of Voicing During Closure, use of full versus partial voicing, and Closure Duration, after controlling for linguistic factors. While speakers differ substantially in the use of each cue, every speaker uses each cue to signal the contrast between voiced and voiceless stops (Figure 2). The plots reveal structured patterns of interspeaker variation, across categories: speakers lie on a continuum of ‘overall’ values of each cue used to signal the stop contrast, including in ‘how voiced’ stops are in general (Figure 4). We also find that speakers differ in the size of the voicing contrast made using VOT and the presence of Voicing During Closure.

Thus, these Scottish English speakers exhibit structured variability in spontaneous speech, for all three cues to stop voicing. Our findings are consistent with Chodroff & Wilson (2018)’s suggestion that the presence of structured variation within phonetic cues acts to help listeners discern linguistic contrasts, whilst other speaker variation facilitates recognition of interspeaker differences in phonetic realization, both personal (Goldinger, 1998; Kleinschmidt, 2019) and systematic, such as social-indexical contrasts (Docherty & Foulkes, 2000). Speakers are both signalling the stop voicing contrast with each of the three cues, and systematically differ from each other with respect to age and time in how they use the cues to realize the contrast.

That we have uncovered such robust evidence for structured variability for each cue, even in this least controlled speech style, is perhaps less surprising than it might seem. Most speech communication takes place in exactly this way; producing words as minimal pairs in citation forms, wordlists or in read sentences, is much less usual. There are hints throughout the literature that the stop voicing contrast in particular is realized in a phonetically less usual way in read citation forms. Lisker & Abramson (1967) specifically note the increased use of voicing lead, and very long VOT, in isolated words, and even more so in minimal pairs, and suggest that this is due to an ‘enhancement effect’. Chodroff & Wilson (2017, p. 41) were surprised by their finding that individual speakers maintain the voicing contrast more in connected speech, where so many other factors can also affect VOT values, than in isolated words. Clayards (2018) also found that producing minimal pairs provoked phonetically prototypical realizations of both stop categories, and interestingly, weaker—rather than stronger—evidence for intraspeaker covariation of multiple cues for stop voicing. Our results for single and multiple cues suggest that we should perhaps take an alternative perspective. We might want to start by assuming that spontaneous speech is the speaker norm, during which talkers need to communicate linguistic and social-indexical contrasts together effectively (Docherty & Foulkes, 2000; Docherty et al., 1997); our findings are consistent with them doing just this. More controlled speech styles, which are crucial for the careful teasing-out of many aspects of phonetic and phonological phenomena, may sometimes end up inhibiting individual speaker behavior.

7.2 Structured variability in multiple phonetic cues for Scottish stop voicing

Three previous studies have shown that speakers exhibit structured variability for multiple cues to stop voicing in English read speech, where these cues are VOT and f_0 at vowel onset (Shultz et al., 2012), plus Closure Duration (Schertz et al., 2015), or following vowel duration (Clayards, 2018). Previous results from experimental work on the three cues considered here (Section 2.1) show the following overall relationships for English: voiceless stops have longer VOT, are less likely to show Voicing During Closure, and have longer Closure Duration; voiced stops have shorter VOT, are more likely to show stop voicing, and have shorter Closure Duration. A pattern similar to a subset of these results also emerged for our spontaneous Scottish English speakers. In Analysis 2, we ran a Canonical Correlation Analysis which identified which pairs of cues are correlated for individual speakers. The two cues which are used most by these individual speakers to realize the voicing contrast, namely VOT and Voicing During Closure, are also significantly correlated: the higher the speaker’s value for VOT for the voicing contrast, the less that speaker will use Voicing During Closure. This means that individual speakers’ use of two cues is coordinated along an axis which is a subspace of the way speakers vary in signaling the voicing contrast, in general: speakers who show more/less aspirated stops will also use Voicing During Closure to a lesser/greater extent. This intra-speaker consistency of phonetic cue coordination seems striking especially because our speech sample is spontaneous vernacular Glaswegian. But, as noted above, perhaps it was easier for us to discern these relationships precisely because our data were more, rather than less, naturalistic (Clayards, 2018). The availability of increasingly large spontaneous datasets means that our assertion can be tested in future work.

7.3 The Scottish stop voicing contrast is changing

Previous work has suggested that Scottish English stops are becoming more aspirated, as part of a more general shift from a more voicing-based system to a more aspiration-based system (Johnston, 1997; Stuart-Smith et al., 2015). The current study provides the first real-time evidence from Scottish spontaneous speech consistent with the assumption that shifts in VOT (reflecting aspiration) are also accompanied by shifts in stop voicing. Analysis 3 considered the evidence for change for each cue across all stops when speakers were grouped by Age and Decade of Birth.

VOT and Voicing During Closure both show evidence consistent with change over time. Inspection of the correlated use of VOT and Voicing During Closure by individual speakers provides greater resolution of the progression of the change in terms of a shift in the trading relation between the cues (cf. Purnell et al., 2005). We find general confirmation of the group results in terms of increasing VOT and reduction of use of Voicing During Closure to make the contrast, except for the 1990s-born adolescents, who revert to more use of Voicing During Closure and less aspirated stops. We also observe that each Age/Decade of Birth group contains both conservative and more innovative speakers with respect to their use of more VOT/less Voicing During Closure. This is exactly the kind of differentiated behavior we expect from members of a community undergoing sound change (Milroy & Milroy, 1985; Stuart-Smith & Timmins, 2010).

Two other factors may be important in shaping our interpretation. The first is SPEECH STYLE. Inference of real-time change depends on the samples from which speakers are drawn. Our sample is drawn from a trend corpus, which includes samples from different speakers

of similar age from the same community recorded at different time. The construction of such a corpus depends on available recordings, and there are numerous likely differences between samples, beyond those relating to language change (Tillery & Bailey, 2003), including recording context.

We cannot rule out the possibility that increase in aspiration/decrease in stop voicing found in the older women results at least partly from style-shifting towards more standard Scottish English in the oral interview context in which they were recorded. However, all but one of the older women, from both time periods, were recorded in interviews, which makes it more difficult to ascribe the differences between the two groups to contextual shifting.

The middle-aged women present a more complex situation: the 1970s recordings are interviews, but those made in the 2000s are casual conversations with a friend, likely to induce more Scots variants, so less aspirated/more voiced stops. But we find the opposite pattern, highly vernacular speech and stops with lengthened VOT and less Voicing During Closure. This supports our inference for a general shift in increased aspiration and reduction of stop voicing after the period of urban regeneration in Glasgow, the mid-1970s and possibly later (cf. Stuart-Smith et al., 2017, 2013).

The second factor is AGE-GRADING. Sociolinguistic accounts of language change (e.g. Sankoff & Blondeau, 2007) note that real-time changes often include patterns which are repeatedly observed as characteristic of a particular age group. For example, many changes show an ‘adolescent peak’, whereby adolescents show increased use of innovative variants compared to younger and older speakers (Tagliamonte & D’Arcy, 2009). Against a trajectory of real-time change towards more aspirated/less voiced stops, the young 1990s-born speakers show a reversal. Rather than a peak of greater aspiration/less voicing, these speakers show an adolescent TROUGH, returning to VOT values typical of the earliest-born speakers, and much more Voicing During Closure.

This pattern of REVERSAL to vernacular variants in adolescents born since Glasgow’s urban regeneration, i.e. during and after the mid-1970s, is found in other variables. For example, young women from the same period show a similar trough in the spectral frequency of /s/, shifting back to productions similar, but different in constriction shape, to women born two generations before (Stuart-Smith, 2019). These speakers are also leading in the adoption of innovative variants such as TH-fronting (Stuart-Smith et al., 2013). Thus this realization of the stop voicing contrast forms part of a wider stylistic construction of sociolinguistic identity (Eckert, 2012), which exploits vernacular features and selected innovations for a ‘new-old Glaswegian’, reminiscent of Watt (2002)’s young Geordies, who reject the ‘flat cap and clogs’ stereotypes of Newcastle, but capitalize on some of the desired characteristics of the ‘toon’, also linguistically. Here, Glasgow vernacular is being restyled and reinvented, combining ideologies of ‘old’, traditional Glasgow (steel and shipyards, classic Billy Connolly), with those of supra-local, post-modern, innovations (cf. Stuart-Smith et al., 2007). Such ‘boomerang’ (Benor, 2015) changes have been observed in other sociolinguistic contexts for other features, including the re-appearance of Yiddish loanwords in American English (Benor, 2015), or the increased stopping of interdental fricatives in Cajun English (Dubois & Horvath, 1998). As here, the enregisterment of the linguistic variation with local identity is a recurring trait for such reversals (cf. Agha, 2003). Their mechanism is unclear, but may be because speakers’ perceptual memories are deeper than their productive repertoires, i.e. they hold exemplars also from their grandparents’ generation (cf. Hay & Foulkes, 2016), which are then available for stylistic exploitation.

Returning to phonology, the evidence for real-time change in our speaker sample is con-

sistent with the assumption that the stop contrast is stable at a structural level, and that its phonetic realization is changing over time, towards greater aspiration and less use of Voicing During Closure. Exactly how to PHONOLOGICALLY characterize this change is an interesting issue. Under a ‘traditional’ view of laryngeal phonology, where both Scots and other English varieties have a [voice] contrast with differing phonetic realization, what we have observed is purely phonetic change, in realization of the [voice] feature. But under the ‘laryngeal realism’ view, where traditional Scots uses [voice] while other English dialects use [spread] to realize the voicing contrast, the change we observe may reflect something deeper—change in the FEATURE carrying the contrast, to [spread]. By the logic of laryngeal realism (e.g. Salmons, 2019), changes in phonetic realization like we observe are a necessary, but not sufficient, condition to demonstrate such a shift. The crucial evidence would come from phonological processes (e.g. assimilation), which is an interesting direction for future work in this or another diachronic corpus of Scots.

8 Conclusion

Weinreich et al. (1968, p. 99) claimed that identifying and accounting for linguistic variation—crucially in its social environment—can lead to a ‘more adequate description of linguistic competence’. The quest to pin down structured heterogeneity has driven the agenda for the last 50 years of variationist sociolinguistics. Purnell et al. (2005) explicitly took this agenda forward, to understand better the fate of stop voicing in Wisconsin English. Their study, on read speech, had a clear diachronic focus. Accounting for the phonetic and linguistic factors which condition speech variation is core to phonetics, and more recently interest has shifted to uncovering ‘structured variability’: how and why speakers differ in how they realize phonological contrasts in non-random ways (Chodroff & Wilson, 2017, 2018). Here we draw together phonetics and variationist sociolinguistics to offer an integrated account of some of the structured heterogeneity that exists for the stop voicing contrast in spontaneous speech from Glaswegian female speakers, over an effective timespan of 100 years.

By working with this stop contrast in its natural habitat, we offer findings together relevant to synchronic and diachronic perspectives on the phenomenon. We show that individual speakers control both linguistic and social-indexical contrasts together. Speakers use all three cues separately, and to an extent together, to signal the stop voicing contrast, but they differ from each other in systematic ways. Most interesting is the multidimensional quasi-synchronic axis along which interspeaker variability is observed, which is at the same time the axis for diachronic change for the same community.

Considering synchronic variability illuminates diachronic change, and vice versa, as predicted in the original manifesto by Weinreich et al. (1968). Moreover, and like Purnell et al. (2005), we see that the change exploits the inherent trading relations for stop voicing. It is also an excellent example of how sound change arises from synchronic variation (Ohala, 1989), though we do not attempt here to specify possible mechanisms.

We also note some caveats and future directions. Our study is necessarily limited. We do not consider other cues to stop voicing, such as f_0 at vowel onset or following vowel duration, which may enhance or change our interpretations. We also require additional social information; here we look only at one ‘corner’ of this dialect. Our description, and inferences, require the inclusion of gender and social class. The former may give insight into the propagation of the change through the community, and the latter is known to exacerbate sociolinguistic polarization and change in this and other British English dialects. Finally, our

assertions concerning the role of speech style on structured variability need testing for this contrast for other dialects, for other languages, within spontaneous speech, and within and across different speech styles from the same speakers. The current availability of high quality acoustic analysis for increasingly large spontaneous speech corpora means that we may not need to wait too long.

References

- ABRAMSON, ARTHUR S., and DOUGLAS H. WHALEN. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63.75–86.
- AGHA, ASIF. 2003. The social life of cultural value. *Language and Communication* 23.231–273.
- AGRESTI, ALAN. 2002. *Categorical data analysis*. 2nd edn. Hoboken, NJ: Wiley.
- AITKEN, ADAM J. 1979. Scottish Speech: A historical view with special reference to the Standard English of Scotland. *Languages of Scotland*, ed. by Adam J. Aitken and Tom McArthur, 85–118. London: Chambers.
- ALLEN, SEAN; JOANNE L. MILLER; and DAVID DESTENO. 2003. Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America* 113.544–552.
- BANG, HYE-YOUNG. 2017. *The structure of multiple cues to stop categorization and its implications for sound change*. McGill University dissertation.
- BARAN, JANE A.; MARSHA Z. LAUFER; and RAY DANILOFF. 1977. Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics* 5.339–350.
- BARR, DALE; ROGER LEVY; CHRISTOPH SCHEEPERS; and HARRY J. TILY. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–278.
- BATES, DOUGLAS; MARTIN MÄCHLER; BEN BOLKER; and STEVE WALKER. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.1–48.
- BENOR, SARAH. 2015. How synagogues became shuls. *Germanic heritage languages in North America*, ed. by Janne B. Johannessen and Joseph C. Salmons, 217–233.
- BOERSMA, PAUL, and DAVID WEENINK. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5.341–345.
- BYRD, DAN. 1993. 54,000 American stops. *UCLA Working Papers in Phonetics* 83.97–116.
- CHEN, MATTHEW. 1970. Vowel length duration as a function of the voicing of the consonant environment. *Phonetica* 22.129–59.
- CHO, TAEHONG; DOUGLAS H. WHALEN; and GERARD DOCHERTY. 2019. Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics* 72.52–65.
- CHODROFF, ELEANOR; JOHN GODFREY; SANJEEV KHUDANPUR; and COLIN WILSON. 2015. Structured variability in acoustic realization: a corpus study of voice onset time in American English stops. *Proceedings of the 18th International Congress of Phonetic Sciences*, ed. by The Scottish Consortium for ICPhS. Glasgow: University of Glasgow. Paper number 632.1–5.
- CHODROFF, ELEANOR, and COLIN WILSON. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61.30–47.
- CHODROFF, ELEANOR, and COLIN WILSON. 2018. Predictability of stop consonant phonetics across talkers: Between-category and within-category dependencies among cues for place and voice. *Linguistics Vanguard* 4.
- CLAYARDS, MEGHAN. 2018. Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica* 75.1–23.
- COLE, JENNIFER; HEEJIN KIM; HANSOOK CHOI; and MARK HASEGAWA-JOHNSON. 2007. Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* 35.180–209.
- CRYSTAL, THOMAS H., and ARTHUR S. HOUSE. 1988. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America* 83.1553–1573.
- DAVIDSON, LISA. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54.35–50. Online: <http://linkinghub.elsevier.com/retrieve/pii/S009544701500073X>.
- DAVIDSON, LISA. 2017. Phonation and laryngeal specification in American English voiceless obstruents. *Journal of the International Phonetic Association* 48.331–356.

- DOCHERTY, GERARD. 1992. *The timing of voicing in British English obstruents*. Berlin/New York: Foris.
- DOCHERTY, GERARD, and PAUL FOULKES. 2000. Speaker, speech and knowledge of sounds. *Phonological knowledge: Conceptual and empirical issues*, ed. by Noel Burton-Roberts, Philip Carr, and Gerard Docherty, 105–129. Oxford: Oxford University Press.
- DOCHERTY, GERARD; PAUL FOULKES; JAMES MILROY; LESLEY MILROY; and DAVID WALSHAW. 1997. Descriptive adequacy in phonology: A variationist perspective. *Journal of Linguistics* 33.275–310.
- DOCHERTY, GERARD; DOMINIC WATT; CARMEN LLAMAS; DAMIEN HALL; and JENNIFER NYCZ. 2011. Variation in voice onset time along the Scottish-English border. *Proceedings of the 17th International Congress of Phonetic Sciences*, ed. by Wai-Sum Lee and Eric Zee, 591–594. Hong Kong: City University of Hong Kong.
- DUBOIS, SYLVIE, and BARBARA M. HORVATH. 1998. Let's tink about dat: Interdental fricatives in Cajun English. *Language Variation and Change* 10.245–261.
- ECKERT, PENELOPE. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41.87–100.
- FRANCIS, ALEXANDER L.; NATALYA KAGANOVICH; and COURTNEY DRISCOLL-HUBER. 2008. Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America* 124.1234–51.
- FROMONT, ROBERT, and JENNIFER HAY. 2012. LaBB-CAT: An annotation store. *Proceedings of the Australasian Language Technology Association Workshop 2012*, 113–117.
- GELMAN, ANDREW, and JENNIFER HILL. 2007. *Data analysis using regression and multi-level/hierarchical models*. Cambridge: Cambridge University Press.
- GOLDINGER, STEPHEN D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105.251–279.
- GONZÁLEZ, IGNACIO; SÉBASTIEN DÉJEAN; PASCAL MARTIN; and ALAIN BACCINI. 2008. CCA: An R Package to extend canonical correlation analysis. *Journal of Statistical Software* 23.1–14. Online: <https://www.jstatsoft.org/v023/i12>.
- GRANT, WILLIAM. 1913. *The pronunciation of English in Scotland*. Cambridge: Cambridge University Press.
- HAY, JENNIFER, and PAUL FOULKES. 2016. The evolution of medial /t/ over real and remembered time. *Language* 92.298–330.
- HONEYBONE, PATRICK. 2005. Diachronic evidence in segmental phonology: The case of obstruent laryngeal specifications. *The internal organization of phonological segments*, ed. by Marc van Oostendorp and Jeroen van de Weijer, 319–354. Berlin: Mouton de Gruyter.
- IVERSON, GREGORY, and JOSEPH SALMONS. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12.369–396.
- IVERSON, GREGORY, and JOSEPH SALMONS. 1999. Glottal spreading bias in Germanic. *Linguistische Berichte* 178.135–151.
- JOHNSTON, PAUL. 1980. A synchronic and historical view of border area bimoraic vowel systems. Edinburgh: University of Edinburgh, MS.
- JOHNSTON, PAUL A. 1997. Regional variation. *The Edinburgh history of the Scots language*, ed. by Charles Jones, 433–513. Edinburgh: Edinburgh University Press.
- KEATING, PATRICIA A. 1984. Phonetic and phonological representation of stop consonant voicing. *Language* 60.286–319.
- KESHET, JOSEPH; MORGAN SONDEREGGER; and THEA KNOWLES. 2014. AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction. Version 0.91. Available at <https://github.com/mlml/autovot/>.
- KESSINGER, RACHEL H., and SHEILA E. BLUMSTEIN. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* 25.143–168.

- KIM, SAHYANG; JISEUNG KIM; and TAEHONG CHO. 2018. Prosodic-structural modulation of stop voicing contrast along the VOT continuum in trochaic and iambic words in American English. *Journal of Phonetics* 71.65–80.
- KLATT, DENNIS H. 1975. Voice onset time, frication and aspiration and aspiration in word-initial consonant clusters. *Journal of Speech, Language and Hearing Research* 18.686–706.
- KLEINSCHMIDT, DAVE F. 2019. Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience* 34.43–68.
- KUZNETSOVA, ALEXANDRA; PER B. BROCKHOFF; and RUNE H. B. CHRISTENSEN. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82.1–26.
- LABOV, WILLIAM. 2001. *Principles of linguistic change. Vol. 2: Social factors*. Oxford: Blackwell.
- LENTH, RUSSELL. 2018. *emmeans: Estimated marginal means, aka least-squares means*. R package version 1.3.1. Online: <https://CRAN.R-project.org/package=emmeans>.
- LIBERMAN, ALVIN M.; PIERRE C. DELATTRE; and FRANKLIN S. COOPER. 1958. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech* 1.153–167.
- LISKER, LEIGH. 1986. Voicing in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29.3–11.
- LISKER, LEIGH, and ARTHUR ABRAMSON. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20.384–422.
- LISKER, LEIGH, and ARTHUR S. ABRAMSON. 1967. Some effects of context on voice onset time in English stops. *Language and Speech* 10.1–28.
- LUCE, PAUL A., and JAN CHARLES-LUCE. 1985. Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *The Journal of the Acoustical Society of America* 78.1949–57.
- MASUYA, YOSHIRO. 1997. Voice onset time of the syllable initial /p, t/ and /k/ followed by an accented vowel in lowland Scottish English. *Phonetics and phonology: Selected papers*, 139–172. Tokyo: Kobian Shobo.
- MILROY, JAMES, and LESLEY MILROY. 1985. *Authority in language: Investigating language prescription and standardisation*. London: Routledge.
- OHALA, JOHN J. 1989. Sound change is drawn from a pool of synchronic variation. *Language change: Contributions to the study of its causes*, ed. by Leiv E. Breivik and Ernst H. Jahr, 173–198. Berlin: Mouton de Gruyter.
- PINHEIRO, JOSE, and DOUGLAS BATES. 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer.
- PURNELL, THOMAS; JOSEPH SALMONS; DILARA TEPELI; and JENNIFER MERCER. 2005. Structured heterogeneity and change in laryngeal phonetics: Upper Midwestern final obstruents. *Journal of English Linguistics* 33.307–338.
- REPP, BRUNO H. 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92.81–110.
- SALMONS, JOSEPH. 2019. Germanic laryngeal phonetics and phonology. *Cambridge handbook of Germanic linguistics*, ed. by Richard Page and Michael T. Putnam. Cambridge: Cambridge University Press. To appear.
- SANKOFF, GILLIAN, and HÉLÈNE BLONDEAU. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83.560–588.
- SCHERTZ, JESSAMYN; TAEHONG CHO; ANDREW LOTTO; and NATASHA WARNER. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52.183–204.
- SCOBIE, JAMES M. 2006. Flexibility in the face of incompatible English VOT systems. *Papers in laboratory phonology 8: Varieties of phonological competence*, ed. by Louis Goldstein, Catherine Best, and Douglas H. Whalen, 367–392. Cambridge: Cambridge University Press.

- SHULTZ, AMANDA A.; ALEXANDER L. FRANCIS; and FERNANDO LLANOS. 2012. Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America* 132.EL95–EL101. Online: <http://asa.scitation.org/doi/10.1121/1.4736711>.
- SONDEREGGER, MORGAN; MAX BANE; and PETER GRAFF. 2017. The medium-term dynamics of accents on reality television. *Language* 93.598–640.
- SONDEREGGER, MORGAN, and JOSEPH KESHET. 2012. Automatic measurement of voice onset time using discriminative structured prediction. *The Journal of the Acoustical Society of America* 132.3965–3979.
- STUART-SMITH, JANE. 2019. Through the looking glass: changing perspectives on /s/ and gender over time in Glasgow. *Linguistics Vanguard* 5, To appear.
- STUART-SMITH, JANE; BRIAN JOSÉ; TAMARA RATHCKE; RACHEL MACDONALD; and ELEANOR LAWSON. 2017. Changing sounds in a changing city: An acoustic phonetic investigation of real-time change over a century of Glaswegian. *Language and a sense of place: Studies in language and region*, ed. by Chris Montgomery and Emma Moore, 38–65. Cambridge: Cambridge University Press.
- STUART-SMITH, JANE, and ELEANOR LAWSON. 2017. Scotland: Glasgow/the Central Belt. *Listening to the past: Audio records of accents of English*, ed. by Ray Hickey, 171–98. Cambridge: Cambridge University Press.
- STUART-SMITH, JANE; GWILYM PRYCE; CLAIRE TIMMINS; and BARRIE GUNTER. 2013. Television can also be a factor in language change: Evidence from an urban dialect. *Language* 89.501–536.
- STUART-SMITH, JANE; MORGAN SONDEREGGER; RACHEL MACDONALD; and TAMARA RATHCKE. 2015. The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology* 33.505–459.
- STUART-SMITH, JANE, and CLAIRE TIMMINS. 2010. The role of the individual in language change. *Language and identities*, ed. by Carmen Llamas and Dominic Watt, 39–54. Edinburgh: Edinburgh University Press.
- STUART-SMITH, JANE; CLAIRE TIMMINS; and FIONA TWEEDIE. 2007. Talkin’ ’Jockney’? Variation and change in Glaswegian accent. *Journal of Sociolinguistics* 11.221–260.
- SUMMERFIELD, QUENTIN. 1975. Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables. *Speech Perception* 2.61–72, Department of Psychology, Queen’s University of Belfast.
- TAGLIAMONTE, SALI A, and ALEXANDRA D’ARCY. 2009. Peaks beyond phonology: Adolescence, incrementation, and language change. *Language* 85.58–108.
- THEODORE, RACHEL M.; JOANNE L. MILLER; and DAVID DESTENO. 2009. Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America* 125.3974–3982.
- TILLERY, JAN, and GUY BAILEY. 2003. Approaches to real time in dialectology and sociolinguistics. *World Englishes* 22.351–65.
- WATT, DOMINIC. 2002. ‘I don’t speak with a Geordie accent, I speak, like, the Northern accent’: Contact-induced levelling in the Tyneside vowel system. *Journal of Sociolinguistics* 6.44–63.
- WEINREICH, URIEL; WILLIAM LABOV; and MARVIN I. HERZOG. 1968. Empirical foundations for a theory of language change. *Directions for historical linguistics*, ed. by Winifred Lehmann and Yakov Malkiel, 95–195. Austin: University of Texas Press.
- WELLS, JOHN C. 1982. *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.
- WESTBURY, JOHN R., and PATRICIA KEATING. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22.145–166.
- YAO, YAO. 2007. Closure duration and VOT of word-initial voiceless plosives in English in spontaneous connected speech. *UC Berkeley Phonology Lab Annual Report*, 183–225.
- YAO, YAO. 2009. Understanding VOT variation in spontaneous speech. *Current numbers in unity and diversity of languages*, ed. by Miok Pak, 1122–1137. Seoul: Linguistic Society of Korea.

ZUE, VICTOR. 1976. *Acoustic characteristics of stop consonants: A controlled study*. Bloomington, IN: Indiana University Linguistics Club.

Notes

¹We do not consider here individual speaker variation in phonetic realization, including change over time, for stop contrasts in languages other than English.

²The larger gap between young and middle-aged speaker groups was partly the result of the recordings available, and partly to ensure clear generational separation between adolescents and their parents.

³With the proviso that Davidson (2016) uses 10–90% as the ‘some voicing’ category, rather than 1–99% as used here.

⁴Fitting a multinomial model requires jointly fitting several component regression models for binary data. Jointly fitting models is computationally difficult and not possible using standard mixed-modeling packages in R (e.g. `lme4`). In a continuation-ratio model, the outcomes of the binary regressions are orthogonal, which means that fitting the binary regressions separately (which is easy in `lme4`) should give equivalent results to fitting them jointly (Agresti, 2002, 7.4, 12.5).

⁵Note that the VOT and CD models only predict log-transformed values, and the VDC models only predict the relative odds of each VDC outcome (**None/Some/All**). To predict differences in ms (VOT, CD) or probabilities (VDC), it is necessary to first fix values of each control predictor. We calculate predictions at each model’s intercept throughout Analysis 1.

⁶About 80% of tokens are post-vocalic (where closure voicing is more likely), while model predictions average over all preceding contexts (vowel, voiced fricative, voiceless fricative), leading to the discrepancy between empirical proportions and model predictions.

⁷For example, the test for ‘overall log(VOT)’ excludes the by-speaker random intercept and its correlation with the by-speaker random VOICING slope.

⁸Note that ‘slope’ is defined differently for VDC than for VOT/CD, so that its expected value is positive for all speakers (higher VDC for voiced stops; higher VOT/CD for voiceless stops).

⁹Note that the models including DECADE OF BIRTH could not themselves be used for Analysis 1, because inclusion of this term means the by-speaker random effects no longer have the interpretation needed for Analysis 1 (each speaker’s characteristic value for the cue’s intercept/slope).