Jane Stuart-Smith*, Morgan Sonderegger, Tamara Rathcke and Rachel Macdonald

The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian

DOI 10.1515/lp-2015-0015

Abstract: While voice onset time (VOT) is known to be sensitive to a range of phonetic and linguistic factors, much less is known about VOT in spontaneous speech, since most studies consider stops in single words, in sentences, and/or in read speech. Scottish English is typically said to show less aspirated voiceless stops than other varieties of English, but there is also variation, ranging from unaspirated stops in vernacular speakers to more aspirated stops in Scottish Standard English; change in the vernacular has also been suggested. This paper presents results from a study which used a fast, semi-automated procedure for analyzing positive VOT, and applied it to stressed syllable-initial stops from a real- and apparent-time corpus of naturally-occurring spontaneous Glaswegian vernacular speech. We confirm significant effects on VOT for place of articulation and local speaking rate, and trends for vowel height and lexical frequency. With respect to time, our results are not consistent with previous work reporting generally shorter VOT in elderly speakers, since our results from models which control for local speech rate show lengthening over real-time in the elderly speakers in our sample. Overall, our findings suggest that VOT in both voiceless and voiced stops is lengthening over the course of the twentieth century in this variety of Scottish English. They also support observations from other studies, both from Scotland and beyond, indicating that gradient shifts along the VOT continuum reflect subtle sociolinguistic control.

***Corresponding author: Jane Stuart-Smith,** English Language/Glasgow University Laboratory of Phonetics (GULP), University of Glasgow, 12 University Gardens, Glasgow G12 8QQ, UK, E-mail: Jane.Stuart-Smith@glasgow.ac.uk
Morgan Sonderegger, Department of Linguistics, McGill University, 1085 Doctor Penfield Avenue, Montreal, QC H3A 1A7, Canada, E-mail: morgan.sonderegger@mcgill.ca
Tamara Rathcke, English Language and Linguistics, School of European Culture and Languages, Cornwallis North West, University of Kent, Canterbury, Kent CT2 7NF, UK, E-mail: T.V.Rathcke@kent.ac.uk
Rachel Macdonald, English Language/Glasgow University Laboratory of Phonetics (GULP), University of Glasgow, 12 University Gardens, Glasgow G12 8QQ, UK, E-mail: rachel.macdonald@glasgow.ac.uk

Keywords: stop consonants, spontaneous speech, automatic measurement, sound change, Scottish English

1 Introduction

Voice onset time (VOT), the time from the burst reflecting stop release until the beginning of quasi-periodicity reflecting the initiation of voicing for a following segment, is long established as a cue to the contrast between voiced and voiceless stops for many languages, including English (e.g., Lisker and Abramson 1964, Lisker and Abramson 1967; Caramazza et al. 1973). VOT may be positive, following the burst, reflecting differing degrees of stop aspiration, or negative, from the onset of voicing during a stop closure until the burst is released, reflecting voicing lead or prevoicing. The behaviour of VOT of stop consonants in varieties of English, and indeed many languages of the world (Cho and Ladefoged 1999), is well known from the numerous studies which swiftly followed the original proposition by Lisker and Abramson (e.g., 1964). In English, voiceless stops tend to show varying degrees of positive VOT, while voiced stops may show much shorter VOT or prevoicing, depending on the presence and/or degree of vocal fold vibration during closure. Our understanding of the factors constraining or promoting variation in VOT is largely based on speech styles which are less usual for most speakers, such as single word elicitation through word lists, read sentences, or read passages. Much less is known about how these factors influence VOT in its more usual habitat, where speakers produce stops most often, i.e., unplanned spontaneous speech (cf. Yao 2009; Sonderegger 2012).

The linguistic context for this study is the vernacular dialect of Glasgow. Scottish English is generally reported to show less aspirated voiceless stops than other varieties of British English, with even less aspiration in vernacular Scots (e.g., Scobbie 2006). But there have also been claims that gradient change towards longer VOT durations more typical of Anglo-English may be underway (e.g., Masuya 1997). The question remains as to whether longer VOT productions for younger speakers in the few recent apparent-time studies of Scottish English demonstrate phonological change in progress, or reflect the results of physiological aging (Docherty et al. 2011). Also, these previous studies of VOT in Scottish English have been based on word lists and read speech.

Here we assess the VOT of stops in age-stratified samples of naturally-occurring spontaneous speech recorded at different time points to gauge whether such patterns are typical across the stylistic repertoires of Scottish English over time. Deriving phonetically robust measures of VOT from

spontaneous speech is more difficult and time-consuming than from read speech or citation forms (Baran et al. 1977). We consider the effects of phonetic and linguistic factors on VOT in these speakers by using an automatic algorithm for detecting positive VOT with manual correction (Sonderegger and Keshet 2012). Our long-term goal is to investigate potential variation and change in the voicing contrast in Scottish English. In this paper we gain an impression of one aspect of the realization of voiced and voiceless stops by considering positive VOT over time.

2 Background

2.1 Variation in VOT: phonetic and linguistic factors

VOT is sensitive to a range of phonetic and linguistic factors, which in turn are subject to language-specific implementation (e.g., Docherty 1992; Cho and Ladefoged 1999; Auzou et al. 2000). Place of stop articulation typically conditions the longest values for velar, and shortest values for bilabial stops (e.g., Cho and Ladefoged 1999). Coronal stops generally show longer VOT durations than bilabials, but may not always be distinct from those of velars: in British English, Docherty (1992) reports a general distinction of bilabial vs non-labial stops; alveolars are not significantly different from velar stops. Following vowel context also conditions VOT. After Lisker and Abramson's (1967) initial negative result for any impact of vocalic environment on VOT, subsequent studies have generally observed longer VOT durations before high close vowels than before low open vowels (e.g., Klatt 1975; Esposito 2002; Berry and Moyle 2011; cf. Mortensen and Tøndering 2013). VOT also varies with speech rate, though differently depending on the voicing of the stop. Specifically, VOT of voiceless stops is negatively correlated with speech rate, whereas for voiced stops there is no correlation or only a small trend (e.g., Summerfield 1975; Miller et al. 1986; Kessinger and Blumstein 1997). A similar asymmetry in the effect of phrasal accent on VOT was found by Cole et al. (2007), with a larger effect for voiceless than for voiced stops.

Other factors considered recently concern aspects of the word in which the stop occurs, position in phrase, and lexical frequency. Cole et al. (2007) anticipated prosodic strengthening of several cues to voicing, including VOT, expecting longer VOT for stops showing phrasal prominence and in phrase-initial position; their examination of read narratives by four American English newscasters found no significant effect of phrase position. Lengthening of VOT was

found in utterance-final position in Yao's (2009) study of unplanned American English spontaneous speech. Yao also found that more frequently used words showed shorter durations (see also Sonderegger 2012 for spontaneous British English speech), though Yu et al.'s (2013) imitation study of single words showed only a non-significant trend in this direction.

2.2 Variation in VOT: social and speaker-specific factors

Several studies have shown that variation in VOT may also be socially conditioned. For example, Ryalls et al. (1997) and Ryalls et al. (2004) considered ethnicity and gender in African-American and Caucasian-American younger male and female speakers (earlier study) and older speakers (later study). Younger speakers showed significant differences in VOT, indicating more voicing of voiced stops in male and African-American speakers; no significant effects of gender or ethnicity were observed in the older speakers. Research on VOT duration and aging does not present straightforward results. Some studies have found that older speakers (e.g., over 70) show shorter VOT durations than younger speakers (e.g., Benjamin 1982; Ryalls et al. 2004), while other studies have either found no significant age-related difference in VOT (e.g., Neiman et al. 1983; Petrosino et al. 1993) or complex interactions between age and gender (Torre and Barlow 2009), suggesting that VOT values may reflect age as a socially-conditioned life stage as much as the results of aging physiology.

Differences in VOT between groups of speakers (e.g., old vs. young) found in studies which do not also control for speech rate may in fact be due to rate (see Morris et al. 2008). However, speech rate is unlikely to explain shorter VOTs in elderly speakers, who typically speak more slowly (e.g., Torre and Barlow 2009). More generally, individual differences in VOT remain even after speaking rate is controlled for (Allen et al. 2003; Yu et al. 2013). Such individual variation is consistent with the idea that VOT can be manipulated as a social-indexical characteristic at the level of the speaker, which may or may not intersect with larger social categories such as age, gender, and ethnicity.

2.3 Variation in VOT beyond read speech

The majority of studies on VOT have used single words elicited through word lists or carrier sentences, as well as reading passages or longer narratives (e.g., Crystal and House 1988; Cole et al. 2007). There has been far less investigation of

unscripted spontaneous speech (cf. Yao 2009; Sonderegger 2012). This is an interesting lacuna because very shortly after their initial exposition of VOT in citation forms in 1964, Lisker and Abramson (1967) wondered about how VOT might vary according to speech style. They found that VOT in read sentences continued to distinguish stops by place of articulation and voicing but differed from isolated-word context in the degree of overlap in distributions between voiced and voiceless stops.

Baran et al. (1977) seem to have been the first to consider VOT of stops in conversational speech. They examined child-directed and adult-directed speech by four American-English speaking mothers, in four styles. They did not find a difference in VOT between child- and adult-directed speech, but they did find that the separation of mean VOT of voiced and voiceless stops was greatest in citation forms (80 ms) and smallest in spontaneous speech (30 ms), due to shorter VOTs for voiceless stops in spontaneous speech (cf. Gósy 2001 for Hungarian; *contra* Krull 1991 for Swedish). Yao (2009) examined VOT for voiceless stops in unplanned spontaneous speech by two American English speakers from the Buckeye Corpus: one male and one female speaker, with particularly fast and slow speaking rates, respectively. VOT was influenced by local speaking rate, place of articulation, lexical frequency, and utterance-final position, though the two speakers did not always show identical patterns. Most recently, Sonderegger (2015) examined VOT in voiceless stops in spontaneous speech by 21 English speakers mostly from across the United Kingdom. VOT was significantly influenced by the same factors, and also syllable stress, following segment type, following vowel height, and speaker gender.

2.4 Short-term shifting in VOT

Variability in VOT is clearly constrained by a complex set of intersecting factors: phonetic, linguistic, prosodic, social, and individual. Changes in aspiration duration in a community's speech over time also presume that variation in VOT during interaction between speakers is accessible to listeners at some level, and is amenable to short-, and longer-term, shifting, as listeners become speakers (cf. Tucker 2007). Evidence that this is the case for short-term shifting is provided by recent research.

Shockley et al. (2004) ran two shadowing experiments, one in which VOT of word-initial English /p t k/ was unaltered, and the second in which VOT was extended. They found that the shadowed speech of both experiments was perceptibly different from baseline reading for listeners performing an

AXB task, and that VOT in the shadowed speech in the second experiment was on average 12 ms longer than that of the baseline. Nielsen (2011) found that speakers lengthened VOT of word-initial /p/ when imitating a set of target words after exposure to productions with artificially-extended VOT. She also found that exposure led to greater VOT in /p/ in novel words, which was also generalized to a new sound (/k/). Imitation was also constrained by lexical frequency and by aspiration duration, as stops with reduced VOTs were not imitated. Yu et al. (2013) explored the impact of manipulating listener attitudes and expectations, as well as personality traits, on speakers' imitations of lengthened VOTs embedded in a narrative. While there was no significant overall change in VOT following exposure (*contra* Nielsen), how much subjects shifted VOT towards or away from the narrator was strongly affected by social and cognitive factors, such as holding a positive attitude towards the narrator. Such studies offer insights into possible mechanisms underlying longer-term change, but are restricted to a couple of time points over a few minutes.

2.5 Longer-term shifting in VOT

Sonderegger (2012) considered day-to-day variability in VOT of 22,500 voiced and voiceless stops in a corpus of spontaneous speech from 12 British contestants on the reality television show *Big Brother UK* over a period of three months. Using regression modelling for voiced and voiceless stops separately, he found that most cases (voiced or voiceless stops, for an individual speaker) showed one or two kinds of change. Daily fluctuation around the mean in VOT was the norm in the data, while about half of cases also showed steady change in a speaker's mean VOT over time. Short-term daily variability in VOT over a timescale of days to months seems to be the norm, and – for some speakers – may lay the foundation for longer-term changes.

Shifts in VOT over a similar timescale have also been observed in bilingual speakers (see, e.g., Flege and Eefting 1987) and are language-specific. Sancier and Fowler (1997) found that a bilingual Brazilian Portuguese/American English speaker showed consistently shorter VOT in Portuguese than in English, but several months' residence in Brazil led to shorter VOT for both languages than a stay of similar length in America. More recently, Balukas and Koops (2014) also showed a language-based asymmetry in voiceless stops in spontaneous codeswitching (New Mexican Spanish/English). Even in long-term language contact situations, long after language acquisition, the language acquired first may continue to exert subtle and consistent effects on VOT.

Very few studies indeed have considered change in VOT over longer timescales.[1] Geiger and Salmons (2006) discuss preliminary results of a small-scale real-time study of aspiration in voiced and voiceless stops in a recessive nineteenth-century German dialect spoken in Wisconsin, which point to a slight reduction in VOT over time, though not in the direction of standard American English. Takada (2012) considered two apparent-time corpora of a large number of Japanese speakers from five regions reading word lists, collected in the late 1980s and the late 2000s. She found indications that the distinctive role of VOT for the voicing contrast is weakening in two regions, though differently in each, even within the same language.

2.6 VOT in Scottish English

Scottish English is reported to have voiceless plosives with less aspiration than in Southern varieties of Anglo-English (Wells 1982; cf. Masuya 1997; Catford 2002). Scottish English comprises a bipolar sociolinguistic continuum of varieties from Scottish Standard English to vernacular Scots (e.g., Aitken 1984). While Scottish Standard English has had less aspirated stops than Anglo-English for some time,[2] syllable-initial stops in Scots are reported as being unaspirated, at least according to commentators writing before the Second World War (Johnston 1997: 505). However, Johnston (1980: 78, in Scobbie 2006: 374) suggests that more aspirated stops are spreading into Scots. Masuya's (1997) small-scale study shows that his 15 Scottish Standard English speakers have shorter mean VOT durations than the five Anglo-English speakers, overall and at each place of articulation (all stops: Scottish, mean = 39.7 ms; Anglo-English, mean = 56 ms), though no statistical tests are given. Masuya's Scottish sample has an apparent-time dimension, with eight speakers born in the 1960s (in their 40s), and the rest born in the 1920s–1940s (in their 60s–80s). Mean VOTs are shorter for the older speakers than for the younger speakers, though there is some overlap. He interprets his results as an indication that aspiration is lengthening in Scottish Standard English, especially in the younger speakers born in the 1960s (see Scobbie 2006).

Differences between degree of aspiration and vernacular/standard accent background in Scottish English are also apparent in Scobbie's (2006) study of word-initial bilabial stops in read wordlists from 12 speakers who were born and raised in

1 There are of course numerous accounts of historical shifts in stop aspiration (e.g., Iverson and Salmons 1995).

2 "When a breathed plosive occurs ... the emission of breath is barely perceptible. It never strikes the ear in the same way as in Southern English or Irish" (Grant 1912: 80).

Shetland, but whose parents fell into three groups in terms of geographical background: Shetland, Scotland, and England. The results showed that individual differences in positive VOT for /p/ could not easily be assigned to a small number of lag categories. Rather there was a range of durations which pattern generally with parental background: informants with vernacular Shetlandic parents show shorter VOTs than those with Scottish parents, though in a gradient fashion. Voiced stops show either prevoicing or short lag VOT, sometimes both in the same speaker. The results are consistent with the assumption of more/less aspirated stops across the poles of the sociolinguistic continuum of Scottish English spoken in Shetland, and also with possibility of ongoing change in VOT.

In a recent and substantial study on VOT in Scottish English, Docherty et al. (2011) analyzed 4,662 tokens of voiced and voiceless plosives, from read wordlists, from 159 speakers in four locations along the Scottish-English border. They found that younger speakers overall showed significantly longer aspiration (and less prevoicing) than older speakers, and attributed this pattern to age grading (older speakers have longer VOT: see Section 2.2) rather than apparent-time change. They also found differences according to location, with VOT shorter for Scottish speakers at the Eastern end (Eyemouth) than for speakers at the Western end of the Border (Gretna), a pattern consistent with their previous findings that Eyemouth speakers use a higher proportion of more 'Scottish' features (e.g., rhoticity, Scottish Vowel Length Rule). As in Scobbie (2006), and in line with the studies of short-term shifts reviewed above (Section 2.3), VOT appears to be subject to subtle sociolinguistic control.

2.7 Research questions for this paper

Previous accounts of Scottish English suggest that two subtle changes may be underway: Scottish Standard English is shifting to longer durations more like Anglo-English (Masuya 1997), and Scots in turn is shifting to durations more like Scottish Standard English (Johnston 1997). At the same time, the evidence to date on Scottish English VOT is restricted to anecdotal observation or measures from single words and read speech collected at a single point in time. Teasing apart age-grading from language change requires inspection of naturally-occurring spontaneous speech from speakers of different ages recorded at different time points. Here we consider stops in spontaneous speech in a vernacular dialect, drawing on the resources of a recently constructed real-time corpus of Glaswegian. While our long-term aim is to investigate potential change in the voicing contrast in this dialect – which would require inspection of positive and negative VOT, as well as other measures capturing voicing during closure and

closure duration – within the scope of this paper, we restrict our focus to a particular dimension, positive VOT, enabling us to observe variability in this particular aspect of the voicing contrast over time. To overcome the time commitment required to obtain large numbers of robust VOT measures from spontaneous speech, we also developed a semi-automated procedure for the task. From this base we address these research questions:

- What factors affect positive VOT in stressed syllable-initial stops in spontaneous Scottish English speech?
- What is the evidence for change in positive VOT over time?

3 Methodology

3.1 Sample

We analyse VOT in stops produced by 23 speakers from the recently created *Sounds of the City* corpus of Glaswegian vernacular. This is a controlled-access, force-aligned, electronic corpus of audio recordings and orthographic transcripts from 142 speakers (around 730,000 words), aligned using LaBB-CAT software (Fromont and Hay 2012). The recordings are of spontaneous speech, and include oral history and sociolinguistic interviews, conversations between friends, and extracts of broadcast speech. The informants are working-class as determined by factors such as socio-economic background, education, and occupation. The corpus is structured by gender, by decade of recording (from the 1970s to the 2000s), and by generation of the speaker (older: 67–90; middle-aged: 40–55; young: 10–15). Its real- and apparent-time structure allows investigation of stability and change effectively across the entire twentieth century. Speech style ranges from very casual to variable style-shifting found in interviews (Johnston 1983); there is also a range in terms of recording quality. Our earlier analysis of 12 speakers of the sample presented here did not show any differences in the effectiveness of our semi-automatic measurement procedure as a result of the type of speech recording (Stuart-Smith et al. 2015). In this study, we did not code or test further for possible additional variation arising from differences in recording context or interlocutor (cf. Tucker 2007).

The sample for this study is shown in Table 1. We worked with the recordings of 23 female speakers, from three age categories, made in the 1970s and the 2000s. The real-time comparison allows us to assess the evidence for change in aspiration over time. The age stratification enables us to consider the influence of physiological age on VOT, specifically whether shorter VOT is found in older speakers, and longer VOT in younger ones. The age stratification also permits apparent-time comparison.

Table 1: Real- and apparent-time dimensions of the sample of 23 speakers from the *Sounds of the City* corpus analysed in this study. 3F = 3 female speakers, and so on.

Real-time	Apparent-time		
Decade of Recording	Old	Middle	Young
1970s	3 F (1890s: Decade of Birth1)	4 F (1920s-b: Decade of Birth3)	4 F (1960s: Decade of Birth5)
2000s	4 F (1920s-a: Decade of Birth2)	4 F (1950s: Decade of Birth4)	4 F (1990s: Decade of Birth6)

This assumes that speakers tend to maintain the pattern of systemic phonetic features which they acquired as children over their lifespan (Sankoff and Blondeau 2007), though it is not yet known how well this assumption holds for VOT, which is demonstrably flexible for some speakers (Sancier and Fowler 1997; Sonderegger 2012). For the apparent-time comparison we would predict longer durations in middle-aged and younger speakers than in older speakers. Style-shifting towards the standard could also induce longer durations, while shifting towards the vernacular would lead to the reverse.

The sample permits comparison by Decade of Recording (70s vs. 00s) and Age (Old vs. Middle vs. Young). Here we compare the six groups as levels of a single factor, Decade of Birth, which enables comparison of each group with each other group in real-time:

- Old: recorded in 70s (born 1890s) vs. recorded in 00s (born 1920s-a)
- Middle: recorded in 70s (born 1920s-b) vs. recorded in 00s (born 1950s)
- Young: recorded in 70s (born 1960s) vs. recorded in 00s (born 1990s)

Additionally, comparison is possible of each group with each other group in apparent-time:

- recorded in the 70s: Old (born 1890s) vs. Middle (born 1920s-b) vs. Young (born 1960s)
- recorded in the 00s: Old (born 1920s-a) vs. Middle (born 1950s) vs. Young (born 1990s)

3.2 Stops

We report the results for singleton voiced and voiceless stops /p t k b d g/ which occurred at the beginning of a stressed syllable (e.g., *people, a'ppear, 'ten,*

a'ttend, etc). Tokens which occurred in words or syllables which were unstressed and/or reduced in the utterance—for example, many that were realized as glottal stops—were excluded. The manual correction of the automatically predicted VOT durations also excluded tokens which were difficult to measure for other reasons, for example, when the burst could not be identified, or when the plosive was strongly lenited or released as a fricative. The procedure for analysing VOT is outlined below in Section 4.

3.3 Linguistic factors

We measured a range of variables for each token that we expected to affect VOT using information from the force-aligned TextGrids, as well as two databases of information about words in British English: CELEX (Baayen et al. 1996) and Subtlex-UK (van Heuven et al. 2014). Variables in italics are included in the models of VOT described below.[3]

- *Place of Articulation* of the stressed-syllable-initial stop was defined as bilabial, alveolar, or velar (3 levels) based on the first pronunciation listed in CELEX.
- *Local Speaking Rate (LSR)* was defined as syllables per second in a phrase, where a phrase was defined as the interval between two intervals of silence of at least 150 ms.
- LSR was used to define two variables included in the models below: its mean value for a given speaker (*Mean Local Speaking Rate*), and the difference between a token's LSR and the mean speaking rate (for the speaker who produced it): the *Speaking Rate Deviation*. This step was taken in view of the substantial variation in how quickly individuals speak, to separate the potential effects of 'faster speakers' (a speaker-level variable) and 'faster speech' (by a given speaker, relative to her average rate) on voice onset time (Theodore et al. 2009).
- *Phrase Position* was defined as initial or medial (2 levels) based on whether the stop occurred at the absolute left edge of a phrase (defined as above).
- *Following Vowel Quality* was defined as high or non-high (2 levels), based on the transcription of the following vowel segment, in turn based on the first pronunciation for the vowel listed in CELEX.

3 In this analysis we did not code or test for position of the stop in the word, i.e., to compare, e.g., /t/ in *tend* vs. *a'ttend*.

– *Word Frequency* (log-transformed) for each token was defined by looking up the orthographic form in Subtlex-UK.[4]

4 Analysis of positive VOT using semi-automatic methods for spontaneous speech corpora

4.1 Positive VOT analysis

It is well known that the voicing contrast of varieties of English cannot be adequately characterized using only VOT (Lisker and Abramson 1967), and certainly not using only positive VOT. We had originally intended to analyse both positive and negative VOT in our dataset, but an interesting anomaly (indeed, result) from this study of spontaneous speech is how voicing is realized in our data, in contrast with previous studies of Scottish English. We found that prevoiced stops with voicing lead, whereby voicing begins at some point during the closure and continues to the burst, were very rare indeed in this dataset (only some 15 instances). Voicing during stop closure tended to appear either as continuous voicing throughout the entire closure, or as no voicing at all; a proportion showed some perseverative voicing into the closure continuing from the preceding voiced segment. The practical outcome for our study was that while the positive VOT algorithm functioned well, the automatic negative VOT algorithm (Henry et al. 2012) was unable to predict negative VOT reliably from these recordings. In ongoing work we have devised other measures for characterizing voicing (e.g., proportion of voicing during closure). We report here only the results for positive VOT.

VOT was annotated by a two-step 'semi-automatic' process: automatic measurement followed by manual correction. Our procedure for measuring positive VOT (for both voiced and voiceless stops) was to identify the period of aperiodic friction following the initiation of a visible burst until the initiation of quasi-periodicity visible from the waveform. This included instances of very short periods of aperiodic friction which occurred after first initial visible spike reflecting the onset of the burst of fully voiced closures (e.g., Nearey and Rochet 1994), though a small degree of damping immediately before release was often observed (this is commonly reported for English, and also Scottish English; see Scobbie 2006: 377–379). This means that our measures of VOT reflect the release phase of

4 One word (Townhead, a place name) was not listed in Subtlex-UK, and the 4 tokens for this word were excluded.

voiceless and voiced stops, including what in previous studies have been counted as long lag ('aspiration'), and short lag and burst duration, respectively.

4.1.1 Step 1: Automatic measurement

We first automatically measured stop VOTs by applying the *AutoVOT* software (Keshet et al. 2014), an implementation of the supervised learning algorithm described in Sonderegger and Keshet (2012). *AutoVOT* uses a set of hand-labelled VOT measurements as a training set of stops to train a structured support-vector machine classifier. Predicting VOT for a new set of stops requires a trained classifier and a window of time in the audio file for each token within which to search for the beginning of the VOT interval. Applying the classifier to each token yields a predicted VOT interval. For this study, one voiceless stop classifier and one voiced stop classifier were trained using around 100 hand-annotated voiceless and voiced stop tokens from (each of) five speakers as training data. The algorithm was then run using these classifiers on the entire recordings of the sample, with search windows determined based on the force-aligned segment boundaries provided by *LaBB-CAT* for each target stop. The algorithm was applied twice, to predict VOT for voiceless and then voiced stop tokens, using the voiceless and then the voiced classifier.

4.1.2 Step 2: Manual correction

Manual inspection, correction, and coding were carried out by four annotators, who were entered into the models as a fixed effect of Annotator. The coding scheme had three labels:

1. The automatic prediction was *Correct*.
2. The automatic prediction was incorrect but easily *Correctable*, and so was corrected.
3. *Not usable*: The stop's location was grossly off due to an alignment error; VOT could not be reliably determined (due to speaker overlap, background noise, or another cause); the token was realized as another segment (fricative, approximant, glottal) or deleted; or there was a transcription error. These tokens were excluded from further analysis.

An annotator could process all instances of voiced or voiceless stops for about 40 minutes of conversational speech in around 40 minutes, sometimes less. This is very much quicker than any process of locating and then hand

labeling stop burst and onset of voicing in spontaneous speech, even using a force-aligned segmentation tier as a guide. The speed of our analysis meant that we could process all possible tokens from each speaker, and so also obtain larger numbers of tokens for analysis.

Predictions were corrected for 5,823 voiced and 4.075 voiceless stops.[5] Table 2 shows the breakdown of tokens by the three labels. 29.8% of voiced and 7.9% of voiceless tokens were coded as *Not usable*. The remaining 4,087 voiced tokens and 3,247 voiceless tokens make up the datasets used to model positive VOT for voiced and voiceless stops presented below.

Table 2: Number and percentage of automatically measured stops by coding label.

	N	*Correct*	*Corrected*	*Not usable*
Voiced stops	5,823	3,171 (54.4%)	916 (15.7%)	1736 (29.8%)
Voiceless stops	4,075	2,689 (76.2%)	558 (15.8%)	828 (7.9%)
All stops	9,898	5,860 (62.6%)	1,474 (15.8%)	2,564 (21.6%)

4.2 Statistical analysis

We modelled VOT as a function of the variables discussed above (Section 3.3), using mixed-effects linear regression models (using the lme4 package in R; Bates et al. 2014). To limit the complexity of the exposition of the results, we fitted separate models for voiceless and voiced stops. Because VOT can only take on positive values in our dataset (Section 4.1), and because the distribution of VOT (for voiceless and voiced stops) is strongly right skewed (Figure 1), we use log (VOT) as the response variable in the models (Sonderegger 2012). We discuss the fixed-effect and random-effect terms included in the models in turn.

4.2.1 Fixed effects

The same eight main effects were included in the voiced and voiceless models:

– the following linguistic factors expected to affect VOT, based on previous work, to address our first research question: properties of the host word (Place of Articulation, Following Vowel Height, Lexical Frequency), of the

5 These counts are *after* excluding 197 voiced and 73 voiceless tokens where there were errors in the manual correction coding or in applying processing scripts.

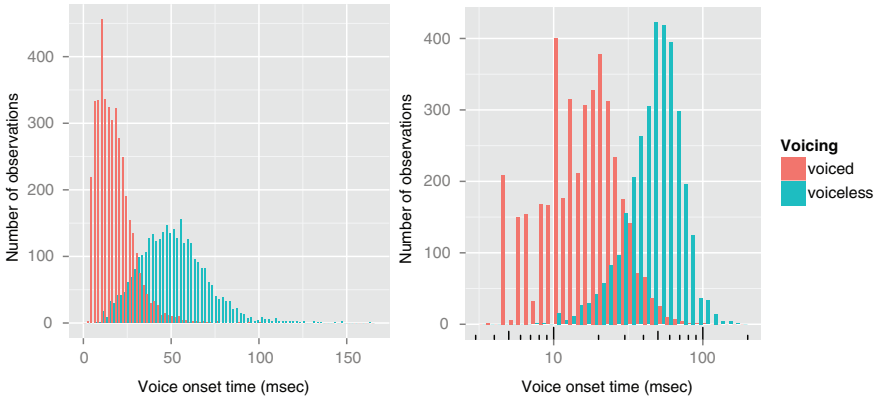

Figure 1: Histogram of VOT for voiced ($n = 4{,}088$) and voiceless ($n = 3{,}247$) stops, untransformed (left) and on a log scale (right).

speaker (Mean Local Speaking Rate), and of the observation (Speaking Rate Deviation, Phrase Position)
- Decade of Birth, to answer our second research question, whether VOT is changing over time (Section 3.1; Table 1)
- Annotator, to account for the possibility that annotators used different criteria in correcting the VOT predictions (Section 4.1)

To facilitate interpretation of the main effect terms in the models and to minimize unnecessary collinearity, categorical variables were coded using Helmert contrasts, with the levels of each variable ordered as follows:[6]
- Place of Articulation: bilabial, alveolar, velar
- Following Vowel Height: low, high
- Phrase Position: initial, medial
- Annotator: 1, 2, 3, 4
- Decade of Birth: 1890s, 1920s-a, 1920s-b, 1950s, 1960s, 1990s (see Section 3.1 above)

The individual fixed-effect coefficients for Annotator were not significant, and so this factor is not discussed further. Continuous variables (Mean Local

6 Helmert contrasts means that the first contrast for Place of Articulation corresponds to 1/2 the difference between alveolar and bilabial (positive = alveolar), the second contrast corresponds to 1/3 the difference between velar and the mean of alveolar and bilabial (0.33*(velar − (alveolar + bilabial)/2)), and so on for other variables. Note that Helmert coding for a factor with two values (such as Phrase Position) is the same as sum coding.

Speaking Rate, Speaking Rate Deviation) were centered (by subtracting the mean), separately within the voiced and voiceless subsets of the data. Main effect terms for the eight variables were included in both models to test hypotheses based on previous work, to test for sources of measurement error, and to address our research questions. To decide which interactions between the eight variables to include in each model, we assessed potential interactions in two ways (separately for the voiced and voiceless data):

1. *Exploratory plots* examining the joint effect of two variables on VOT in the empirical data (such as Figure 6). Pairs of variables in which one variable seemed to modulate the other variable's effect on VOT were flagged as potential interactions.
2. *Stepwise backwards model selection*,[7] beginning with a model with random intercepts only (by-speaker and by-word), and all possible two-way interactions between the eight variables, with the exception of Annotator (since this variable was included only as a control for overall inter-annotator differences). Terms were dropped using an alpha = 0.01 significance level, due to the large number of comparisons being performed.

Interactions that were selected by both methods were included as fixed effects: for voiced stops, the interactions between Place of Articulation and Decade of Birth, and between Place of Articulation and Speaking Rate Deviation; for voiceless stops, the interactions between Place of Articulation and Decade of Birth, between Speaking Rate Deviation and Decade of Birth, and between Frequency and Decade of Birth.

4.2.2 Random effects

By-word and by-speaker random effects were included to account for the fact that tokens from individual words (voiced: 376 levels; voiceless: 550 levels) and speakers (voiced and voiceless: 23 levels) are not independent. By-word and by-speaker random intercepts were included to account for differences in VOT among speakers and words, after controlling for other sources of variability (Allen et al. 2003; Sonderegger 2012). All possible by-word and by-speaker random slopes were included in each model, to account for variability among speakers and words in the influences on VOT captured by the fixed-effect terms (such as speaking rate: Theodore et al. 2009), and to guard against Type I error in the fixed-effect coefficients (Barr et al. 2013). However, correlations between

7 Performed using step( ) in the lmerTest package in R (Kuznetsova et al. 2014).

random-effect terms were not included, since doing so led to unidentifiable models. To a certain extent, including by-speaker random intercepts and slopes also controls for additional situational factors such as recording context on VOT, which were not included as fixed factors in the models.

4.2.3 Procedure and diagnostics

After fitting initial models for the voiced and voiceless data with the fixed and random-effect terms described above, examination of the residuals showed that they were mostly normally-distributed, with the exception of a small fraction of tokens (about 1%) far from the origin which caused the residual distributions to be skewed. Since these points are likely to have an undue influence on the model fits, points with residuals more than 3 SD from the origin were excluded (voiced: 28 points; voiceless: 35 points) (Baayen 2008). The models were then refitted to the trimmed datasets, with the result that the residual distributions were brought closer to normality. It is the results of these models that are reported below.

The condition number of the model matrix was 6.8 for voiced stops and 7.3 for voiceless stops, indicating a low level of collinearity between predictors, unlikely to affect model results (Belsley et al. 1980; Baayen 2008).[8] The (Pearson) correlation between fitted values and log(VOT) was $r = 0.693$ for the voiced model and $r = 0.733$ for the voiceless model ($r^2 = 0.480$, 0.537). Thus, the models explain approximately 48% and 54% of variability in VOT for the voiced and voiceless stops.

5 Results

We now present the model results with respect to our two research questions. We focus first on those factors which affect VOT independently of a speaker's age and date of recording (Sections 5.1 and 5.2). Then in Section 5.3 we consider the evidence for whether the stop contrast may be changing over time (terms involving Decade of Birth). Full statements of the results are given in Tables 5–7 in the Appendix. Table 5 presents the Type 3 analysis of variance (ANOVA) table for the fixed effects included in each model, with denominator degrees of freedom, *F*-value, and corresponding *p*-value calculated using Satterthwaite's approximation

8 Besley et al. (1980: 105) characterize kappa of 5–10 as indicating 'weak' collinearity.

(using the lmerTest package in R). Tables 6 and 7 summarize the fixed-effect coefficients for each model. Coefficient significances were computed using *t*-tests with degrees of freedom computed using the Satterthwaite approximation (again using lmerTest). The random effect variances are given in Table 8.

5.1 Word-level variables

We first consider variables defined at the level of the word: the stop's voicing (i.e., voiced vs. voiceless), its place of articulation, the following vowel's height, and word frequency.

The empirical distribution of VOT for voiced and voiceless stops in Figure 1 clearly shows that the voicing contrast is maintained through positive VOT for these speakers. This is confirmed by comparing the predicted estimates from the two models. Exponentiating the estimated intercepts for the voiceless and voiced models gives predicted VOTs of 46.5 ms and 15.5 ms, when all other predictors are held at their average values.[9] The 99% confidence intervals for these intercepts (using a Wald test) are [42.2, 51.3] ms and [13.6, 17.6] ms.

In both the voiceless and voiced models, place of articulation significantly affects VOT (Place of Articulation: voiced: $F(2,23.4) = 83.8$; $p < 0.0001$; voiceless: $F(2,28.3) = 42.5$, $p < 0.0001$). Due to the presence of an interaction of Place of Articulation with Decade of Birth in both models (Section 5.3), these main effects can be interpreted as showing that VOT does differ significantly by place of articulation, averaging over all groups of speakers. To get a better sense of how place affects VOT, post-hoc Tukey tests were carried out for Place of Articulation for each model. For voiced stops, we find the commonly found pattern of bilabial < alveolar < velar (/b/ < /d/, /g/; /d/ < /g/: $p < 0.0001$); Cho and Ladefoged 1999). For voiceless stops, bilabials had lower VOT than alveolars, which did not differ significantly from velars (/p/ < /t/, /k/: $p < 0.0001$; /t/ = /k/: $p = 0.28$; see Docherty 1992).[10] These patterns are reflected in the empirical distribution of VOT by place of articulation shown in Figure 2.

9 More precisely: because all categorical predictors in the voiceless and voiced VOT models have been Helmert-coded and all continuous predictors were centered, the intercept can be interpreted as the predicted value of the response (log(VOT)) when continuous predictors are held at their average values, averaged across predictions for all levels of each categorical variable.

10 Tukey post-hoc tests were carried out using ghlt in the multcomp package in R (Hothorn et al. 2008), adjusting for multiple comparisons using the single-step method, and averaging over interactions with Place of Articulation and over covariates.

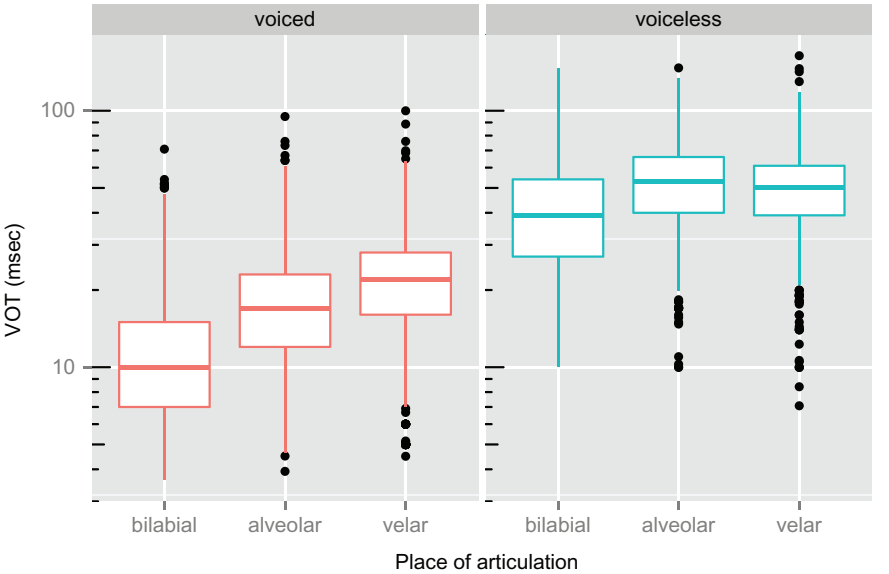

Figure 2: Boxplots of log(VOT) by Place of Articulation, for voiced (left: $n = 4{,}088$) and voiceless (right: $n = 3{,}247$) stops.

The effects of following vowel height and word frequency on VOT for both voiceless and voiced stops are in the expected directions (longer VOT before high vowels than before non-high vowels; shorter VOT in more frequent words), but do not reach significance (Following Vowel Height: voiced $p = 0.15$, voiceless $p = 0.55$; Lexical Frequency: voiced $p = 0.94$, voiceless $p = 0.21$).

5.2 Speaker-level and observation-level variables

We now consider the influence of variables describing properties of speakers (except Decade of Birth; see Section 5.3 below) and observations: Annotator, Mean Speaking Rate, Speaking Rate Deviation, and Phrase Position.

For both voiceless and voiced stops, which annotator corrected the VOT predictions for a given speaker does not significantly affect VOT (Annotator: voiced: $F(3,12.4) = 1.40$, $p = 0.29$; voiceless: $F(3,12.3) = 0.86$, $p = 0.49$). This gives confidence in the quality of the semi-automatic measurement process, and suggests that annotators used very similar criteria in correcting the automatic VOT measurements.

A speaker's mean local speaking rate did not significantly affect VOT for either voiceless or voiced stops (Mean Local Speaking Rate: voiced $p = 0.92$, voiceless

$p = 0.76$), although in both cases the effect is in the expected direction (VOT decreases for faster mean speaking rate). On the other hand, speaking rate relative to the speaker's mean does affect VOT (Speaking Rate Deviation: voiceless $\hat{\beta} = -0.022,\ p = 0.012;$ voiced $\hat{\beta} = -0.016$, $p = 0.052$), although only marginally for voiced stops, with VOT decreasing for faster speech. However, the effect both has a larger effect size and is more significant for voiceless than for voiced stops, reflecting the pattern seen in empirical plots of VOT as a function of speaking rate deviation (Figure 3). These differences between the voiceless and voiced stop speaking rate effects are in line with previous work on global speaking rate effects in lab speech (e.g., Miller et al. 1986; Kessinger and Blumstein 1997).

Phrase-medial stops have lower VOT than phrase-initial stops, as anticipated. The effect has a larger effect size and is much more significant for voiced than for voiceless stops, with voiceless stops only reaching marginal significance (Phrase Position: voiceless $\hat{\beta} = -0.025,\ p = 0.092;$ voiced $\hat{\beta} = -0.046$, $p = 0.0040$). Having said that, it is clear from the empirical plots of VOT versus phrase position in Figure 4, that phrase position has only a very small effect on VOT relative to other variables.

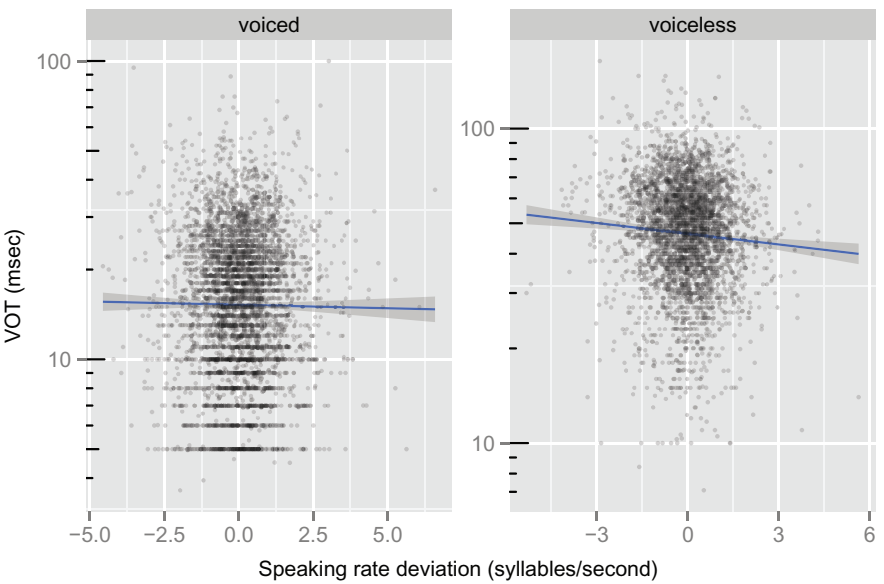

Figure 3: Scatterplot of log(VOT) and speaking rate deviation (difference between local speaking rate and a speaker's mean speaking rate), for voiced (left: $n = 4{,}088$) and voiceless (right: $n = 3{,}247$) stops, with a linear smoother superimposed (solid line; shading represents 95% confidence intervals).

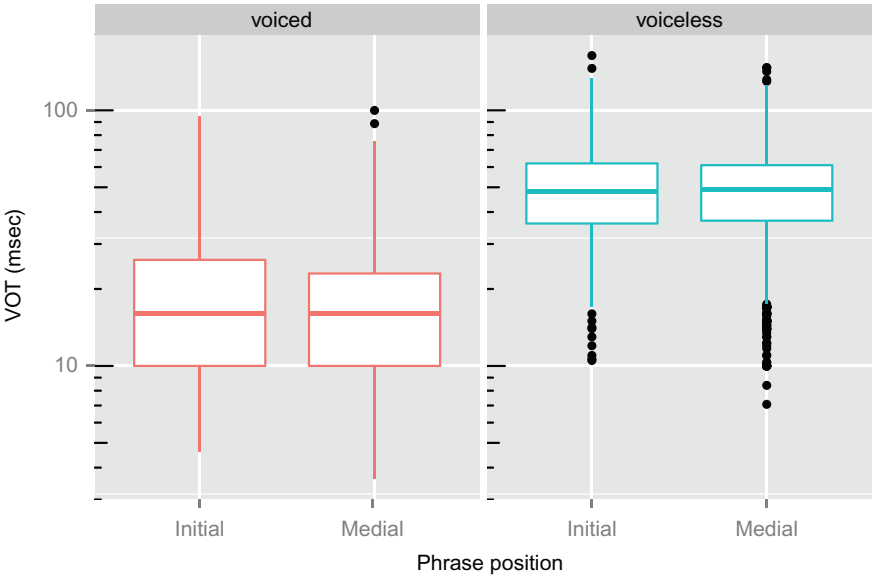

Figure 4: Boxplots of log(VOT) by phrase position, for voiced (left: $n = 4{,}088$) and voiceless (right: $n = 3{,}247$) stops.

5.3 Variables relating to time

We now turn to the models' results related to our second research question: whether VOT is changing over time. As discussed in Section 3.1 above, the sample was structured to consider evidence of change from two perspectives:

- *Real-time change*: differences in VOT in speakers of the same age group (Old, Middle, Young) between the two decades of recording (1970s vs. 2000s)
- *Apparent-time change*: differences in VOT between speakers of different age groups (Old vs. Middle vs. Young) within the same decade of recording.

We examine the two models' predictions for these two types of change, which involves asking if VOT differs between nine pairs of level comparisons of Decade of Birth: three real-time comparisons (1970s Old vs. 2000s Old, etc.) and six apparent-time comparisons (1970s Old vs. 1970s Middle vs. 1970s Young, etc.).

That is, we make nine pairwise comparisons for a factor with six levels. Because these comparisons are not independent, we must control for multiple comparisons. At the same time, there is significant debate about exactly how and whether to correct for multiple comparisons in mixed models (Gelman et al. 2012).

Thus, in all results presented below where we examine the model's predictions for both real time and apparent time, we present both uncorrected p-values (p) and p-values corrected using the Bonferroni method (p_{corr}), a conservative method of adjusting for multiple comparisons. These can be thought of as minimally and maximally conservative p values, with the 'real' value falling somewhere in between. Given that we assume that we may be witnessing a subtle effect, and that our modelling is maximally conservative with the inclusion of both random intercepts and slopes, and the most conservative correction for comparison, we consider here the results in full, both the small number of significant effects and the numerical tendencies.

To assess whether 'overall' change has occurred – the most straightforward interpretation of our second research question – we consider the main effect of Decade of Birth, averaging across other variables. However, the presence of significant interactions with Decade of Birth suggests that the main effects alone may not tell the whole story. To assess whether change has occurred for some types of words and not others, we examine in more detail the interaction of Decade of Birth with Place of Articulation, which appears in empirical plots to be important to take into consideration in assessing change in VOT for both voiced and voiceless stops (see Figure 6 below). We also briefly discuss other interactions with Decade of Birth in the models, as well as an interaction between variables other than Decade of Birth. Again, we note that our modelling of the possible impact of time on VOT in this dataset is statistically very conservative, and we therefore give both uncorrected and corrected p-values when assessing real-time and apparent-time change.

5.3.1 Main effects

The effect of decade of birth on VOT is very marginal for voiced stops, and just significant for voiceless stops (Decade of Birth: voiced $F(5,12.9) = 1.78$, $p = 0.19$; voiceless $F(5,12.7) = 3.15$, $p = 0.045$). Thus, there is weak evidence that VOT shows "overall" dependence on when a speaker was born, i.e., averaging across variables involved in interactions with Decade of Birth, as is evidenced in the empirical distribution of VOT by decade of birth (Figure 5). A trend visible in this data is that VOT tends to increase as a function of decade of birth, in both real and apparent time, provided that the final group of speakers born in the 1990s are left out; these speakers tend to have lower VOT than any other group. Because of the presence of interactions with Decade of Birth in the models, we do not conduct post-hoc tests here to see if these trends in 'overall' VOT are

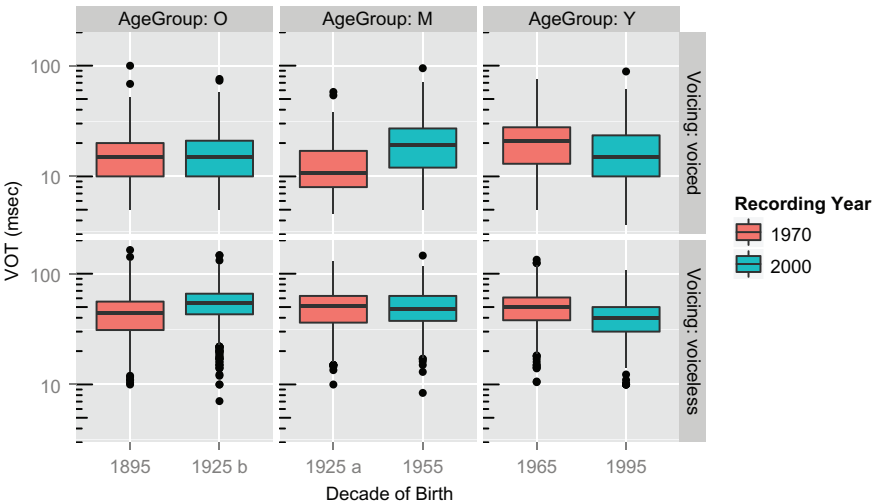

Figure 5: Boxplots of log(VOT) by Decade of Birth showing real-time comparisons, for voiced ($n = 4{,}088$) and voiceless ($n = 3{,}247$) stops, for Old, Middle, and Young speakers.

borne out statistically, and instead turn to interpreting change in VOT in the presence of these interactions.

5.3.2 Interactions of Decade of Birth with Place of Articulation

We interpret the effect of Decade of Birth primarily by checking for real-time and apparent-time change in log(VOT) *within* each Place of Articulation. Figures 6 and 7 show the empirical distribution of VOT by Decade of Birth and Place of Articulation. As we consider any possible evidence for real-time and apparent-time change, it is useful to refer to these figures to understand the models' predictions. Real-time comparisons correspond to comparing the left and right parts of a panel. For example, the upper left-hand panel of Figure 7 compares VOT for /p/ tokens for Old speakers recorded in the 1970s and the 2000s. Apparent-time comparisons correspond to comparing boxes for the same recording year on a given row. For example, the red boxes in the top row of Figure 7 compare VOT for /p/ tokens across the three age groups of speakers in the 1970s.

Recall that checking for real- and apparent-time change involves making nine comparisons, so that doing so for all three places of articulation for voiced and voiceless stops entails 54 comparisons ($9 \times 3 \times 2$). To simplify the presentation

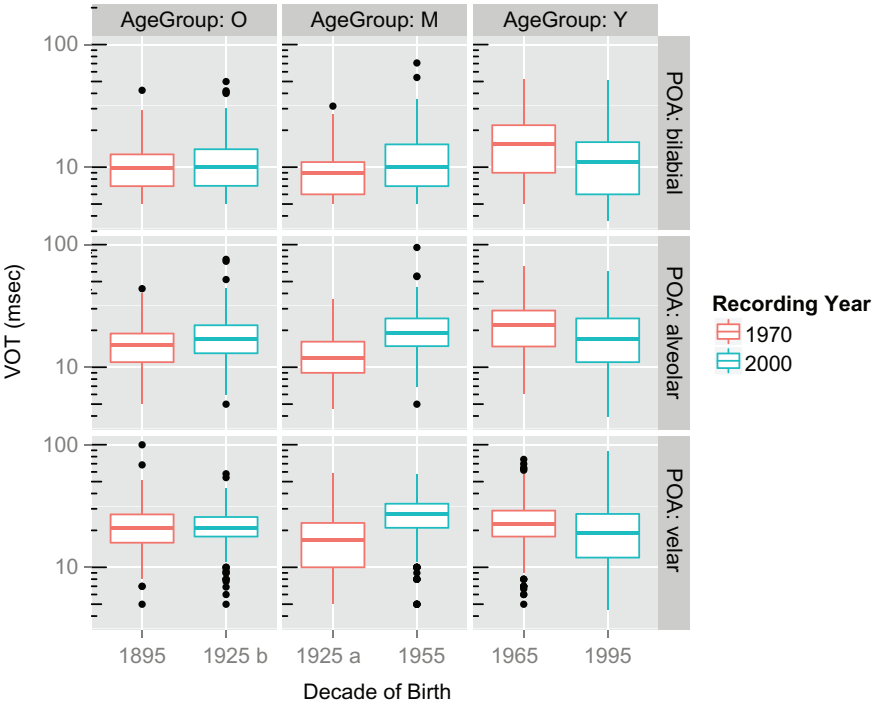

Figure 6: Boxplots of log(VOT) by Decade of Birth and Place of Articulation, for voiced stops ($n = 4{,}088$). Real-time comparisons are between 1970s and 2000s decade of recording, for the same age group. Apparent-time comparisons are between different age groups for the same decade of recording. O = Old; M = Middle; Y = Young.

of this large number of comparisons, we consider real-time and apparent-time change in turn.

5.3.2.1 Real-time change

To consider the evidence for real-time change in VOT for stops at each place of articulation, we estimated the difference in log(VOT) between the 1970s and 2000s decade of recording for each of the three age groups, for the voiceless and voiced stop models. These estimated differences are presented in Table 3, with significances computed via *t* tests with degrees of freedom based on the Satterthwaite approximation.[11]

11 All estimated differences and associated statistics (*t*, *df*, *p*) were calculated using the lsmeans package in R (Lenth 2014).

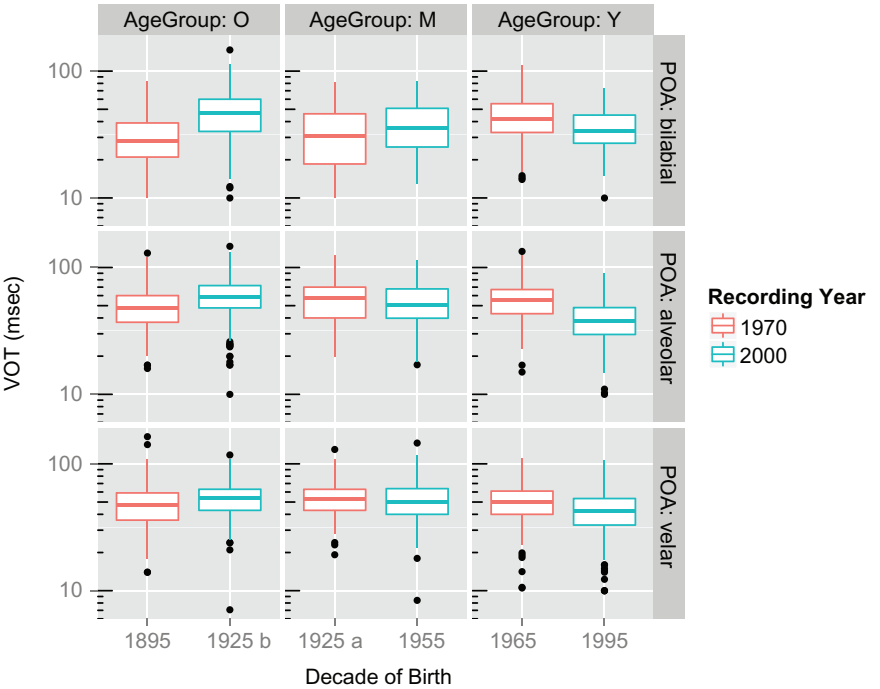

Figure 7: Boxplots of log(VOT) by Decade of Birth and Place of Articulation, for voiceless stops ($n = 3{,}247$). Real-time comparisons are between 1970s and 2000s decade of recording, for the same age group. Apparent-time comparisons are between different age groups for the same decade of recording. O = Old; M = Middle; Y = Young.

The first observation to make based on these estimated differences is that, taking a maximally conservative statistical approach, this dataset offers modest evidence for real-time change in VOT, since most do not reach significance, even using uncorrected *p*-values. Nonetheless, the differences which reach significance at an alpha = 0.05 level (uncorrected), bolded in Table 3, suggest what such a change might look like. For voiced stops, VOT increases for /d/ and /g/ for middle-aged speakers from the 1970s to the 2000s (/d/: est. diff. = 0.427, $p = 0.033$; /g/: est. diff. = 0.484, $p = 0.019$). For voiceless stops, VOT increases for /p/ and /t/ for old speakers from the 1970s to the 2000s (/p/: est. diff. = 0.557, $p = 0.0020$; /t/: est. diff. = 0.345, $p = 0.0388$). All these increases in VOT are clearly visible in the empirical data (Figure 6: middle column, bottom two panels; Figure 7: left column, top two panels). Thus, the significant differences are consistent with the inference of a lengthening of VOT in real time from the 1970s to the 2000s.

Table 3: Real-time comparisons based on the models for voiced and voiceless stops: estimated differences in log(VOT) between 1970s and 2000s, for each age group (Old, Middle, Young), within each place of articulation (bilabial, alveolar, velar). Each estimated difference is shown with its associated standard error, *t* statistic, and uncorrected and corrected significance. Estimated log(VOT) differences which reach significance at the 0.05 level (uncorrected *p*-values), along with the corresponding significances, are bolded. Positive estimated log(VOT) differences are italicized.

Age group	Place of articulation	Estimated difference	Std. Err	*df*	*t*	*p*	p_{corr}
Voiced stops							
Old	Bilabial	*0.181*	0.209	17.86	0.866	0.3981	1.0
	Alveolar	*0.293*	0.208	17.6	1.41	0.1758	1.0
	Velar	*0.208*	0.213	18.78	0.977	0.3409	1.0
Middle	Bilabial	*0.037*	0.188	21.18	0.197	0.8459	1.0
	Alveolar	***0.427***	0.186	20.31	2.293	**0.0326**	0.88
	Velar	***0.484***	0.189	21.04	2.553	**0.0185**	0.58
Young	Bilabial	−0.133	0.249	17.06	−0.534	0.6001	1.0
	Alveolar	−0.207	0.250	17.14	−0.829	0.4186	1.0
	Velar	−0.187	0.255	18.41	−0.731	0.4737	1.0
Voiceless stops							
Old	Bilabial	***0.557***	0.158	20.7	3.539	**0.0020**	0.054
	Alveolar	***0.345***	0.156	19.91	2.212	**0.0388**	1.0
	Velar	*0.219*	0.149	17.02	1.47	0.1598	1.0
Middle	Bilabial	*0.128*	0.146	27.35	0.874	0.3897	1.0
	Alveolar	*0.046*	0.142	23.72	0.323	0.7494	1.0
	Velar	*0.036*	0.133	19.48	0.269	0.7910	1.0
Young	Bilabial	−0.225	0.186	19.02	−1.211	0.2407	1.0
	Alveolar	−0.280	0.183	17.97	−1.531	0.1433	1.0
	Velar	−0.131	0.177	15.96	−0.74	0.4701	1.0

This interpretation is bolstered if we set aside which estimated differences are significantly different from zero, and simply examine the set of estimated differences in log(VOT) between the 1970s and 2000s in Table 3, together with the empirical data in Figure 6 and Figure 7. Two clear patterns are apparent from the estimated means. First, and perhaps surprisingly, speakers born in the 1990s (decade of recording = 2000s, age group = Y) have *lower* VOT than other groups, resulting in every estimated difference involving this group being negative. Second, considering only speakers born in other decades, there is a perfect pattern of VOT increasing between the 1970s and 2000s decades of recording for old and middle-aged speakers, across places of articulation, for both voiced

and voiceless stops, though sometimes by only a small amount. Both of these patterns are largely borne out in the empirical data. Anticipating our discussion in Section 6, our interpretation of the results for change over time is that VOT is moderately increasing over time for both voiced and voiceless stops, but that speakers born in the 1990s have unusually low VOTs perhaps reflecting a shift to vernacular norms which is consistent with other aspects of their stylistic repertoire.[12] For now, we note that the models' results (significant and tendencies) regarding real-time change provide modest but consistent evidence for this conclusion.

5.3.2.2 Apparent-time change

To test for apparent-time change in VOT for stops at each place of articulation, we estimated the difference in log(VOT) between each pair of age groups (Old vs. Middle, Middle vs. Young, Old vs. Young) within each of the 1970s and 2000s decades of recording, for the voiceless and voiced stop models. These estimated differences (with associated *p*-values, etc.), calculated as for the real-time change comparisons, are presented in Table 4.

Table 4: Apparent-time comparisons based on the models for voiced and voiceless stops: estimated differences in log(VOT) within the 1970s and 2000s decades of recording, between each pair of age groups (Young, Middle, Old), within each place of articulation (bilabial, alveolar, velar). Each estimated difference is shown with its associated standard error, *t* statistic, and uncorrected and corrected significance. Estimated log(VOT) differences which reach significance at the 0.05 level (uncorrected *p*-values), along with the corresponding significances, are bolded. Positive estimated log(VOT) differences are italicized.

Comparison	Place of articulation	Estimated difference	Std. Err	*df*	*t*	*p*	p_{corr}
1970s sample (voiced stops)							
Mid v. Old	Bilabial	*0.196*	0.251	16.62	0.782	0.445	1.0
	Alveolar	*0.094*	0.250	16.27	0.377	0.711	1.0
	Velar	*0.015*	0.254	17.09	0.058	0.955	1.0
Young v. Mid	Bilabial	*0.207*	0.177	23.99	1.171	0.253	1.0
	Alveolar	*0.322*	0.176	23.46	1.834	0.079	1.0
	Velar	*0.250*	0.181	25.23	1.381	0.180	1.0

(*continued*)

12 There may well be other factors leading to the reduced VOTs in these younger speakers which result from social, stylistic, and/or situational factors; the impact of discourse context factors is being pursued in ongoing work.

Table 4: (*continued*)

Comparison	Place of articulation	Estimated difference	Std. Err	*df*	*t*	*p*	p_{corr}
Young v. Old	Bilabial	*0.403*	0.262	16.35	1.538	0.143	1.0
	Alveolar	*0.416*	0.262	16.41	1.586	0.132	1.0
	Velar	*0.265*	0.266	16.99	0.997	0.333	1.0
2000 sample (voiced stops)							
Mid v. Old	Bilabial	*0.053*	0.226	16.8	0.233	0.8185	1.0
	Alveolar	*0.228*	0.225	16.56	1.012	0.3259	1.0
	Velar	*0.290*	0.227	16.9	1.278	0.2184	1.0
Young v. Mid	Bilabial	*0.037*	0.213	18.45	0.172	0.8655	1.0
	Alveolar	−0.312	0.214	18.76	−1.455	0.1621	1.0
	Velar	−0.420	0.219	20.05	−1.918	0.0694	1.0
Young v. Old	Bilabial	0.089	0.318	14.77	0.281	0.7829	1.0
	Alveolar	−0.084	0.319	14.89	−0.264	0.7954	1.0
	Velar	−0.130	0.323	15.58	−0.403	0.6927	1.0
1970 sample (voiceless stops)							
Mid v. Old	Bilabial	*0.261*	0.191	20.26	1.364	0.1874	1.0
	Alveolar	*0.185*	0.187	18.74	0.988	0.3356	1.0
	Velar	*0.178*	0.182	16.69	0.98	0.3411	1.0
Young v. Mid	Bilabial	*0.264*	0.135	28.06	1.953	0.0609	1.0
	Alveolar	*0.017*	0.130	24.44	0.133	0.8953	1.0
	Velar	−0.076	0.124	21.05	−0.613	0.5464	1.0
Young v. Old	Bilabial	***0.524***	0.194	17.86	2.7	**0.0147**	0.40
	Alveolar	*0.203*	0.193	17.35	1.051	0.3075	1.0
	Velar	*0.102*	0.188	15.66	0.545	0.5937	1.0
2000 sample (voiceless stops)							
Mid v. Old	Bilabial	−0.169	0.169	19.18	−0.995	0.3319	1.0
	Alveolar	−0.114	0.168	18.34	−0.68	0.5047	1.0
	Velar	−0.006	0.161	15.75	−0.035	0.9728	1.0
Young v. Mid	Bilabial	−0.090	0.163	22.31	−0.551	0.5873	1.0
	Alveolar	−0.309	0.161	21.34	−1.916	0.0689	1.0
	Velar	−0.243	0.151	17.1	−1.604	0.1269	1.0
Young v. Old	Bilabial	−0.258	0.236	16.25	−1.096	0.2889	1.0
	Alveolar	−0.423	0.234	15.69	−1.811	0.0893	1.0
	Velar	−0.249	0.229	14.56	−1.085	0.2955	1.0

The main observation to make from these estimated differences is again that under this conservative statistical modelling strategy, our dataset offers very modest evidence in terms of significant effects for apparent-time change in VOT. The one significant result is that VOT is larger for /p/ for young speakers than for old speakers in the 1970s decade of recording (est. diff = 0.524, $p = 0.0147$), reflecting the pattern observed in the empirical data (Figure 7: top left and top right panels). This is consistent with the interpretation introduced above, of an increase in VOT over time, leaving aside speakers born in the 1990s.

As for real-time change, we can gain additional insight by also examining the set of estimated apparent-time differences in VOT in Table 4 with the empirical data shown in Figures 6 and 7. For apparent-time comparisons for voiced and voiceless stops, we see that for the 1970s recordings, both Middle-aged and Younger speakers show longer estimates than Old and Middle-aged speakers, respectively (all bar one instance where Younger speakers show very slightly longer /k/ than Middle-aged speakers). The pattern is similar for the 2000s recordings for voiced stops, except that, as expected, the Young speakers born in the 1990s show generally shorter estimates. For voiceless stops occurring in the 2000s recordings, the Younger speakers again show shorter estimates; so too do the Middle-aged speakers, though these are very short.

6 Discussion

We have presented the results of an investigation into positive VOT of stops in a real- and apparent-time sample of naturally-occurring spontaneous speech from female speakers of Glaswegian vernacular. Our study was motivated by two reasons. First, there is rather little information about VOT in spontaneous – and vernacular – speech, despite the fact that VOT is otherwise an extremely commonly investigated aspect of stop production. Second, the few existing findings on Scottish English have suggested a possible lengthening in progress over the twentieth century, especially for voiceless stops. To be able to carry out a feasible study, we also wanted to develop a fast, reliable method of measuring VOT from casual spontaneous speech. We structure the discussion of our results by considering the evidence bearing on our two research questions, but begin by considering our methodology.

6.1 Methodology – using *AutoVOT* to measure VOT in spontaneous speech

We developed a semi-automatic procedure to process large numbers of reliable VOT measures, using the *Auto-VOT* algorithm developed by Sonderegger and Keshet (2012). It was trained on an initial set of some 500 hand-measured tokens, applied to the force-aligned data, and then the algorithm's VOT predictions when applied to the full dataset were manually checked. This method worked well because it yielded large numbers of phonetically good VOT measures much faster than our previous experiences of analyzing aspects of natural speech from sociolinguistic corpora. Also, we were pleased to find no statistical effect of individual annotator. Four different phoneticians acted as annotators, including one who joined the study after almost half of the speakers' data had been corrected. This shows that, using our method, swift correction can easily be transferred to a new annotator, without introducing bias into the analysis.

The overall performance of *Auto-VOT* on the Glaswegian corpus was good and similar to the results presented in Sonderegger and Keshet (2012), who carried out an evaluation of the performance of the algorithm on four different datasets comparing it to that of human transcribers for the same data. The results for the two datasets closest to our sample are the *Switchboard* corpus of American speech and the *Big Brother UK* dataset of spontaneous British speech; for both corpora, VOT detection windows were placed *manually*, rather than using force-aligned segment boundaries. The proportion of VOT measures that agree to within 10 ms between independent human transcribers is 70% and 74% for the two datasets, and 68% and 74% for the comparison of a human transcriber with *Auto-VOT*.

Our diagnostic of performance is slightly different because we have no data transcribed by multiple human transcribers, so we consider the proportions judged by the annotators to be *Correct* or *Correctable*. For this dataset we found that for voiced and voiceless stops taken together, *Auto-VOT* predicted 63% of VOT measures which were *Correct*. This increases to 76% for voiceless stops, and is lower for voiced stops at 54%. If we include those stops which were coded as *Correctable*, this gives 78% for all stops, 92% for voiceless stops and 70% for voiced stops. These performance rates are impressive in comparison to the *Switchboard* and *Big Brother UK* datasets, given that neither suffered from incorrect force-alignment boundaries or issues of poor recording quality, nor from frequent lenited realizations of plosives, which were common for our Glaswegian speakers. Clearly our

method is optimal for voiceless stops, since much of the data can be measured accurately, with very little data loss. Our finding that VOT in voiced stops is more difficult to measure than in voiceless stops, especially in spontaneous speech, was also found by Baran et al. (1977), who reported similar proportions of 'measurable' stops: 75% for voiceless stops versus 51% for voiced ones. Overall, we feel that our semi-automated procedure using *Auto-VOT* is very promising for future phonetic analyses of VOT in stops in naturally-occurring speech.

6.2 VOT in spontaneous Scottish vernacular speech

Our first research question led us to investigate the evidence for previously-observed constraints on positive VOT for stops in spontaneous Scottish English. Our main findings are that the expected effects of phonetic and linguistic factors are generally also observed here, and that there are both similarities and differences in the observations relating to VOT in Scottish English stops according to speech style.

6.2.1 Phonetic and linguistic factors and VOT in spontaneous Scottish vernacular

Place of articulation exerted constraints on VOT in our vernacular Glaswegian data similar to those that have been found elsewhere, though we note a difference according to voicing. Voiced stops showed the hierarchy often reported, with increasing VOT from bilabial < alveolar < velar (Lisker and Abramson 1964; Cho and Ladefoged 1999). For voiceless stops bilabials had shorter VOT than both alveolars and velars, which were not significantly different from each other, as has been found for other speakers of British English (Docherty 1992).

VOT has been observed to increase as overall global speech rate decreases for voiceless stops, but to show no significant effect of speaking rate for voiced stops (e.g., Miller et al. 1986; Kessinger and Blumstein 1997). We found a similar pattern, with significantly shorter VOT for faster local speaking rate for voiceless stops and marginally shorter VOT for voiced stops. This pattern was not found for *mean* local speaking rate, which gives a speaker's mean rate across all pause-bounded phrases, but for the local speaking rate *deviation*, which gives the difference of a token's speaking rate (i.e., the phrase) from the speaker's mean rate. Thus, it is only the latter, speaker-independent measure of speaking rate

that is significantly correlated with VOT.[13] Our finding of a weak but marginally significant relationship for voiced stops, in contrast to previous work, may be due to our consideration of more data than previous laboratory studies. Or it may be because a larger range of speaking rates was elicited in our conversational speech sample, which would present another advantage of examining the effects of prosodic factors on VOT in spontaneous speech.

The usefulness of spontaneous speech data for investigating prosodic factors is also indicated by our finding for phrase position. Cole et al. (2007) anticipate prosodic strengthening to be reflected in a range of acoustic measures, including increased VOT, for accented stops and stops in phrase-initial position. They found no significant prosodic strengthening effects, perhaps because of the nature of the speech style or smaller numbers of tokens. We found evidence for a very slight but significant increase in VOT in voiced stops, and a marginal increase in voiceless stops, for phrase-initial position in comparison to phrase-medial ones. Exploratory plots examining VOT as a function of distance from the right edge of the pause-bounded phrase did not show any relationship (*contra* Yao 2009).

We also expected to find effects of vowel height and lexical frequency on VOT, but these only appeared as non-significant tendencies, albeit in the right direction for both voiced and voiceless stops. VOT was longer before high vowels (Klatt 1975; Berry and Moyle 2011), and shorter in more frequent words (Yao 2009; Sonderegger 2012). Previous studies have also reported similar tendencies; we suspect that significant effects may be detected with a larger sample size (see, e.g., Lisker and Abramson 1967 on vowel height, and Yu et al. 2013 on lexical frequency).

6.2.2 VOT in spontaneous Scottish English speech

Our estimates for the entire speaker sample, irrespective of Decade of Birth, for voiceless and voiced stops, predicted from the regression model and so taking into account all factors included in the model, are 46.5 ms and 15.5 ms, respectively (38 ms for /p/ and 11.3 ms for /b/). Masuya (1997) reports an overall VOT mean of 39.7 ms for all voiceless stops for his Scottish Standard English speakers, and 33.3 ms for those born before the 1950s. Scobbie (2006) gives 56 ms for /p/ and 15 ms for /b/ for all his speakers, and 34.8 ms/14.6 ms for those

13 This measure is more comparable with previous work where generally only one 'speaking rate' measure is used, but speakers are *asked* to speak at fast or slow speeds, i.e., relative to their mean rate.

with vernacular Shetlandic parents. Absolute comparison between our study and these is tricky given numerous differences, especially in speech style and gender (both studies included male informants), but what is immediately striking is that VOT from phrase-medial words in read sentences (Masuya) and phrase-initial words in word lists (Scobbie) looks shorter or about the same as those from spontaneous speech (our data). Given that we would expect VOT in spontaneous speech to be *shorter* (cf., e.g., Baran et al. 1977), this may perhaps be another indication that VOT is lengthening over time in Scottish English, though further direct comparative work is needed to investigate the impact of speech style on VOT.

We can only make cautious statements about the general nature of the voicing contrast based on our results, since we only considered positive VOT for the voiced stops (which included all aspects of the release phase, burst, and aspiration), as we were unable to measure voicing lead in terms of negative VOT in our dataset. However, as noted before, this first reservation is also a finding. The realization of voicing in voiced stops in citation forms and read speech seems to be rather different from that in spontaneous speech, because the incidence of prevoicing, even of absolute phrase-initial voiced stops, was so rare. Around 10% of our data were phrase-initial, but we identified only some 15 instances of prevoicing in the entire dataset (i.e., less than 0.3%). Here voiced stops had either entirely voiced closures, or no voicing at all during closure, with only a small proportion of closures showing some perseverative voicing. Admittedly, we could have hand-measured the duration of fully voiced closures and reported these as negative VOT durations. However, more generally, the treatment of voicing during closure in spontaneous speech is complex in spontaneous speech, and we are addressing this in ongoing work. What we do observe from positive VOT is clear separation of voiced and voiceless stops in terms of predicted estimates (Section 5.1), with some overlap in the distributions of both raw and log(VOT) measures, as shown in Figure 1 (cf. Lisker and Abramson 1967; Baran et al. 1977).

6.3 Real-time change in VOT in Scottish English

Our second research question asks whether there is evidence consistent with change over time in VOT in this dataset of spontaneous Glaswegian. We consider our results in terms of the effects of age on VOT and the evidence for inferring change from the real- and apparent-time comparisons.

Docherty et al. (2011) found that, overall, younger speakers showed longer VOT than older speakers, but they were reluctant to interpret these findings as indicating change over time, given evidence from some previous work that VOT decreases over the lifespan (e.g., Benjamin 1982). Masuya (1997) was convinced that his apparent-time results should be interpreted in terms of change, but his older speakers were all over 60 when his recordings were made. So these too could be the results of physiological age differences. Only Johnston's (1997) comment about more aspirated stops being found in Scots does not depend on apparent time data.

We found a rather different pattern. Our elderly speakers recorded in the 1970s show significantly shorter VOT for /p/ than younger speakers recorded in the same decade; tendencies in the same direction are found for /b d g/ and /t k/, again in the same decade. But for speakers recorded in the 2000s, younger speakers show consistent tendencies for shorter VOT than elderly speakers. Also, elderly speakers recorded in the 2000s show significantly longer VOT durations for /p t/ than for those recorded in the 1970s, as well as tendencies in the same direction for /k/.

These results suggest that VOT reflects more than physiological age for these speakers of Glaswegian. On the one hand, VOT durations can clearly be manipulated independently of age, because the youngest speakers, adolescents, show the shortest VOT. On the other, we appear to be witnessing lengthening of VOT over time even in our oldest speakers (aged from 67 to 90), who would be expected to show the shortest VOTs (e.g., Benjamin 1982). Our results are more like those of Torre and Barlow (2009), which suggest that VOT in their speakers reflects local socially-determined categories of social age and gender, since in their study older men had the shortest VOT of all groups, but their older women had the same durations as younger women. Other phonetic features are known to be constrained by physiological factors, but can also be manipulated according to factors of social identity. For example, the peak frequency of /s/ in Glaswegian is influenced by the sex of the speaker, with males generally showing lower frequency /s/ than female speakers, but also affected by social gender, with middle-class females and working-class women showing high-frequency /s/ but working-class girls having the same frequency /s/ as male speakers (Stuart-Smith 2007). Physiological age may lead to reduced VOT (for reasons which are not yet clear), but other social factors operating in a community can also condition shorter or longer durations, depending on the specific social meanings conveyed through stop release and aspiration (cf. Podesva et al. 2015).

So it seems as if our VOT data are not age-graded, or at least not in the direction predicted by physiological age. If we consider the results from our

conservative statistical analyses of both real and apparent time, we have a few significant results supported by tendencies in the same direction, which are consistent with the assumption that VOT in both voiced and voiceless stops is lengthening, and indeed has lengthened, over the course of the twentieth century. Specifically, we find that middle-aged speakers show longer VOT for /b g/ over real time, and older speakers show longer VOT for /p t/; younger speakers show longer VOT for /p/ than older speakers recorded in the 1970s. This evidence is in line with Masuya's (1997) apparent-time data, albeit for Scottish Standard English. It also confirms Johnston's (1997) observation about Scots (though without particular evidence that this shifting relates to the standard). However, this finding has to be qualified by the reservation that it applies only to the speakers recorded in the 1970s, and the old and middle-aged speakers recorded in the 2000s. The younger speakers recorded in the 2000s conversely show tendencies for shorter VOT in both real and apparent time.

Two observations are necessary. The first relates to speech style. The recordings in our corpus are not all of the same nature. Some are interviews, while others are casual conversations between friends. All of the old speakers were recorded in interviews, whereas our middle-aged and young samples contain both interviews and conversations. Scottish Standard English is thought to have longer aspiration than Scots vernacular, especially for voiceless stops (Johnston 1997; Scobbie 2006). It is possible that some of the lengthening we observe in our old and middle-aged speakers relates not only to change over time but may also be the result of style-shifting towards longer durations typical of Standard Scottish English. However, the middle-aged women recorded in the 2000s were talking in casual conversations with close friends, while those in the 1970s participated in sociolinguistic interviews with a university fieldworker. The significant real-time result is that the more recently recorded women, also born more recently, show longer VOT for voiced stops than those recorded and born earlier, even though stylistically we would predict shorter VOT in the more casual style, and – if speakers were accommodating to the standard – longer VOT in the interview style. While it is never possible to disentangle the effects of style and time in these data, these patterns suggest that VOT may be lengthening over time in vernacular Scots.

The second point relates to the behaviour of the younger speakers recorded in the 2000s, who show consistent real- and apparent-time tendencies for shorter VOTs than all other speakers. Why should this group be using less aspirated stops in contrast to the general trend towards lengthening of VOT? Previous sociolinguistic research on working-class adolescents in

Glasgow has shown that, since the 1990s, this group of speakers strongly orient to non-standard vernacular norms for a range of other phonetic and phonological features. For example, adolescents recorded in the 1990s use more vernacular lexical variants (e.g., *h*[ʉ]*s* for *house*) than adolescents recorded in the 1970s (Stuart-Smith 2003). They also use more instances of Scots vocalized /l/, e.g. *a, ba*, for *all, ball*̩ than Macafee observed in the 1980s (Stuart-Smith et al. 2006). This shift towards non-standard variation, including the rapid adoption of non-local variants (e.g., TH-fronting; Stuart-Smith et al. 2013), appears to be part of a more general sociolinguistic polarization within the city between working-class and middle-class adolescents (Stuart-Smith et al. 2007). The appearance of stops with shorter VOT durations, more typical of vernacular Scots (even when change is in progress for this variety) looks congruent with such a shift away from lengthened tokens, especially if the lengthening is associated with Standard Scottish English.[14] At the same time, given the observed flexibility of VOT with numerous social, stylistic, situational, and even cognitive factors (e.g., Nielsen 2011; Yu et al. 2013), further work is also needed to try to discover whether there are other aspects of these speakers' spontaneous recordings which may also contribute to their much shorter VOT durations.

7 Conclusions

VOT is an aspect of stop production which has been intensively examined, and yet surprisingly little work has considered it in stops when they occur in their most usual environment, naturally-occurring spontaneous speech (Yao 2009; Sonderegger 2012). This is at least partly because measuring VOT in conversational speech can be difficult and time consuming (Baran et al. 1977). Here we considered VOT in a vernacular dialect where change in progress has been mooted (e.g., Johnston 1997), but for which only a few studies of read speech exist, and confounds of the influence of physiological age on VOT occur with those of possible inference of (apparent-time) change (Docherty et al. 2011). We used a semi-automatic procedure based on Sonderegger and Keshet's (2012) Auto-VOT algorithm for predicting positive VOT, which requires some training data and the existence of boundaries roughly indicating the beginning of the stop. Using a fast manual coding scheme to correct the predictions of Auto-

14 This would also argue against influence from exposure to Anglo-English long lag stops via the broadcast media for these adolescent speakers.

VOT, we gained reliable VOT measures for over 7,000 stops from 23 female speakers of Glaswegian stratified by age (decade of birth) and decade of recording.

Regression modelling confirmed expected constraints on VOT in spontaneous speech for place of articulation of the stop and the speaker's speech rate, and showed some evidence for prosodic strengthening in slightly increased VOT in phrase-initial stops. Vowel height and lexical frequency were not significant but showed expected tendencies. We did not find that VOT was always shorter in our oldest speakers. Rather our conservative statistical treatment of the data showed consistent tendencies, with a few statistically significant instances, for the inference of real- and apparent-time lengthening of VOT in voiced and voiceless stops in all groups of speakers bar those born in the 1990s and recorded in the 2000s, who show shorter VOT durations, even controlling for speech rate. This last group has also been observed to be reverting to more non-standard vernacular norms, both local and non-local, for other phonological features (e.g., Stuart-Smith et al. 2007). The use of stops with shorter VOT, more usually associated with vernacular Scots, may be part of a more general construction of local, non-standard sociolinguistic personae.

Our study of VOT in stops in spontaneous speech offers a real- and apparent-time perspective on a range of factors which have been shown in laboratory studies of read speech together to constrain the patterning of VOT, from phonetic and linguistic factors to those which relate to local social-indexical meanings as reflected by patterns consistent with variation and change over time. We suspect that our results may also herald a more fundamental shift in the phonetic realization of this contrast over time for Scots vernacular, but this remains the subject of future work.

Acknowledgments: This paper substantially extends an earlier study which gave preliminary results for 12 speakers (Stuart-Smith et al. 2015). JSS is very grateful to the Leverhulme Trust for the funding which supported this research (RPG-142). MS was supported by the Social Sciences and Humanities Research Council (430-2014-00018) and the Fonds de recherche du Québec – Société et culture (183356). We are very grateful to Cordula Klein for help with manual correction of the TextGrids, to Misha Schwartz for scripts to parse the TextGrids, and to Thea Knowles for help with *AutoVOT* and comments on drafts. Audiences at ICLaVE7, NWAV34, and LabPhon14 gave us valuable feedback on earlier versions. We are also grateful to the two anonymous reviewers and to the editors of this special issue of *Laboratory Phonology*.

Appendix A

Table 5: Analysis of variance table for fixed effects in models of log(VOT) for voiced and voiceless stops, with *F* statistic, denominator degrees of freedom, and *p*-values calculated using Satterthwaite's approximation.

Predictor	Sum Sq	Mean Sq	NumDF	DenDF	*F*	*p*
Voiced stops						
FOLLOWING VOWEL HEIGHT	0.0012	0.0012	1	53.62	2.11	0.15
PLACE OF ARTICULATION (POA)	30.48	15.24	2	23.43	83.76	<**0.0001**
DECADE OF BIRTH	1.44	0.29	5	12.94	1.78	0.19
PHRASE POSITION	2.18	2.18	1	16.62	11.16	**0.0040**
SPEAKING RATE DEVIATION	0.72	0.72	1	18.07	4.35	0.051
MEAN LOCAL SPEAKING RATE	0.07	0.07	1	12.37	0.011	0.92
ANNOTATOR	0.70	0.23	3	12.42	1.4	0.29
FREQUENCY	0.0041	0.0041	1	40.49	0.006	0.94
POA:DECADE OF BIRTH	2.89	0.29	10	16.03	1.61	0.19
POA:SPEAKING RATE DEVIATION	1.62	0.81	2	39.62	4.68	**0.015**
Voiceless stops						
FOLLOWING VOWEL HEIGHT	0.012	0.012	1	64.54	0.359	0.55
PLACE OF ARTICULATION (POA)	6.76	3.38	2	28.39	42.562	<**0.0001**
DECADE OF BIRTH	1.42	0.28	5	12.72	3.145	**0.046**
PHRASE POSITION	0.37	0.37	1	20.43	3.117	0.092
SPEAKING RATE DEVIATION	0.72	0.72	1	19.91	7.556	**0.012**
MEAN LOCAL SPEAKING RATE	0.022	0.022	1	12.39	0.092	0.77
ANNOTATOR	0.24	0.079	3	12.34	0.86	0.49
FREQUENCY	0.15	0.15	1	104.96	1.60	0.21
SPEAKING RATE DEV.:DECADE OF BIRTH	1.23	0.25	5	17.40	2.67	0.058
FREQUENCY:DECADE OF BIRTH	0.88	0.18	5	15.90	2.21	0.10
POA:DECADE OF BIRTH	2.02	0.20	10	15.99	2.33	0.063

Table 6: Summary of fixed effects for the model of log(VOT) for voiceless stops: coefficient estimates ($\hat{\beta}$), standard errors, associated *t*-statistics, and significances. Significances below the alpha = 0.05 level are bolded. Subscripted predictors correspond to contrasts of categorical variables. (See text.)

Predictor	$\hat{\beta}$	$SE(\hat{\beta})$	*df*	*t*	*p*
Intercept	−3.07	0.038	17.9	−81.68	<**0.0001**
1. Word-level variables					
FOLLOWING VOWEL HEIGHT	0.0097	0.016	64.5	0.60	0.55
PLACE OF ARTICULATION$_1$	0.15	0.022	22.7	6.50	<**0.0001**
PLACE OF ARTICULATION$_2$	0.068	0.010	38.5	6.61	<**0.0001**
FREQUENCY	−0.0068	0.084	105.0	−1.26	0.21
2. Speaker-level variables					
DECADE OF BIRTH$_1$	0.105	0.084	12.43	1.242	0.24
DECADE OF BIRTH$_2$	0.090	0.031	12.19	2.886	**0.014**
DECADE OF BIRTH$_3$	0.022	0.037	12.46	0.592	0.56
DECADE OF BIRTH$_4$	0.013	0.019	12.34	0.671	0.51
DECADE OF BIRTH$_5$	−0.027	0.031	12.53	−0.875	0.40
MEAN LOCAL SPEAKING RATE	−0.031	0.103	12.39	−0.30	0.77
ANNOTATOR$_1$	0.061	0.051	12.58	1.20	0.25
ANNOTATOR$_2$	0.020	0.036	12.02	0.56	0.59
ANNOTATOR$_3$	−0.007	0.020	12.36	−0.38	0.71
3. Observation-level variables					
SPEAKING RATE DEVIATION	−0.022	0.008	19.91	−2.75	**0.012**
PHRASE POSITION	−0.025	0.014	20.43	−1.77	0.092
4. Interactions					
POA$_1$:DECADE OF BIRTH$_1$	−0.019	0.038	16.05	−0.50	0.62
POA$_2$:DECADE OF BIRTH$_2$	−0.007	0.017	24.44	−0.45	0.66
POA$_1$:DECADE OF BIRTH$_3$	−0.029	0.019	11.87	−1.54	0.15
POA$_2$:DECADE OF BIRTH$_4$	−0.023	0.008	15.06	−2.90	**0.011**
POA$_1$:DECADE OF BIRTH$_5$	−0.008	0.014	14.32	−0.55	0.59
POA$_2$:DECADE OF BIRTH$_1$	0.000	0.006	17.60	−0.05	0.96
POA$_1$:DECADE OF BIRTH$_2$	−0.021	0.010	11.96	−2.06	0.062
POA$_2$:DECADE OF BIRTH$_3$	−0.011	0.004	16.41	−2.54	**0.022**
POA$_1$:DECADE OF BIRTH$_4$	−0.019	0.009	17.30	−2.03	0.058
POA$_2$:DECADE OF BIRTH$_5$	−0.001	0.004	24.90	−0.18	0.86
DECADE OF BIRTH$_1$:SPEAKING RATE DEV.	0.014	0.017	31.04	0.78	0.44
DECADE OF BIRTH$_2$:SPEAKING RATE DEV.	−0.008	0.007	18.04	−1.06	0.31
DECADE OF BIRTH$_3$:SPEAKING RATE DEV.	0.012	0.005	18.83	2.22	**0.039**
DECADE OF BIRTH$_4$:SPEAKING RATE DEV.	0.007	0.004	16.43	1.78	0.094
DECADE OF BIRTH$_5$:SPEAKING RATE DEV.	0.004	0.003	14.54	1.32	0.21
DECADE OF BIRTH$_1$:FREQUENCY	0.016	0.007	22.99	2.27	**0.033**
DECADE OF BIRTH$_2$:FREQUENCY	0.003	0.003	11.91	1.01	0.33
DECADE OF BIRTH$_3$:FREQUENCY	0.002	0.002	17.05	0.81	0.43
DECADE OF BIRTH$_4$:FREQUENCY	0.002	0.002	12.17	1.34	0.20
DECADE OF BIRTH$_5$:FREQUENCY	0.004	0.002	36.79	2.00	0.053

Table 7: Summary of fixed effects for the model of log(VOT) for voiced stops: coefficient estimates ($\hat{\beta}$), standard errors, associated *t*-statistics, and significances. Subscripted predictors correspond to contrasts of categorical variables. Significances below the alpha = 0.05 level are bolded. (See text.)

Predictor	$\hat{\beta}$	SE($\hat{\beta}$)	*df*	*t*	*p*
Intercept	−4.170	0.050	15	−83.192	<**0.0001**
1. Word-level variables					
FOLLOWING VOWEL HEIGHT	0.022	0.015	54	1.452	0.15225
PLACE OF ARTICULATION$_1$	0.192	0.023	23	8.4	<**0.0001**
PLACE OF ARTICULATION$_2$	0.145	0.015	23	9.781	<**0.0001**
FREQUENCY	0.00039	0.005	41	−0.078	0.94
2. Speaker-level variables					
DECADE OF BIRTH$_1$	0.051	0.116	12	0.438	0.67
DECADE OF BIRTH$_2$	0.059	0.043	13	1.36	0.20
DECADE OF BIRTH$_3$	0.077	0.051	13	1.522	0.15
DECADE OF BIRTH$_4$	0.035	0.027	13	1.317	0.21
DECADE OF BIRTH$_5$	−0.006	0.043	13	−0.14	0.89
MEAN LOCAL SPEAKING RATE	−0.015	0.143	13	−0.105	0.92
ANNOTATOR$_1$	−0.082	0.070	13	−1.174	0.26
ANNOTATOR$_2$	0.036	0.050	12	0.72	0.49
ANNOTATOR$_3$	−0.044	0.028	12	−1.602	0.13
3. Observation-level variables					
SPEAKING RATE DEVIATION	−0.016	0.008	18	−2.085	0.051
PHRASE POSITION	−0.046	0.014	17	−3.34	**0.0040**
4. Interactions					
POA$_1$:DECADE OF BIRTH$_1$	−0.026	0.038	16	−0.67	0.51
POA$_2$:DECADE OF BIRTH$_2$	−0.022	0.024	14	−0.911	0.38
POA$_1$:DECADE OF BIRTH$_3$	0.027	0.020	14	1.37	0.19
POA$_2$:DECADE OF BIRTH$_4$	0.004	0.013	13	0.321	0.75
POA$_1$:DECADE OF BIRTH$_5$	0.036	0.014	16	2.459	**0.026**
POA$_2$:DECADE OF BIRTH$_1$	0.015	0.009	13	1.647	0.12
POA$_1$:DECADE OF BIRTH$_2$	−0.006	0.012	21	−0.52	0.61
POA$_2$:DECADE OF BIRTH$_3$	−0.009	0.007	17	−1.219	0.24
POA$_1$:DECADE OF BIRTH$_4$	−0.010	0.010	22	−1.037	0.31
POA$_2$:DECADE OF BIRTH$_5$	−0.007	0.006	21	−1.075	0.29
POA$_1$:SPEAKING RATE DEVIATION	0.0044	0.0071	3205	0.62	0.53
POA$_2$:SPEAKING RATE DEVIATION	−0.0135	0.0045	21	−3.00	**0.0069**

Table 8: Estimated variances and corresponding standard deviations for random-effect terms in the model of log(VOT) for voiced stops.

Group	Variable	Est. variance	Est. SD
Speaker	INTERCEPT	0.037	0.19
	FOLLOWING VOWEL HEIGHT	0.00059	0.024
	PLACE OF ARTICULATION$_1$	0.0068	0.082
	PLACE OF ARTICULATION$_2$	0.0027	0.052
	SPEAKING RATE DEVIATION	0.00046	0.022
	PHRASE POSITION	0.000069	0.026
	FREQUENCY	0.000064	0.0080
	POA$_1$:SPEAKING RATE DEVIATION	0	0
	POA$_2$:SPEAKING RATE DEVIATION	0.000029	0.0053
Word	INTERCEPT	0.0083	0.091
	PHRASE POSITION	0	0
	MEAN SPEAKING RATE	0.00036	0.019
	DECADE OF BIRTH$_1$	8.10E-08	0.00028
	DECADE OF BIRTH$_2$	0.00093	0.030
	DECADE OF BIRTH$_3$	0.00017	0.013
	DECADE OF BIRTH$_4$	0.0011	0.033
	DECADE OF BIRTH$_5$	0.00099	0.0314

Table 9: Estimated variances and corresponding standard deviations for random-effect terms in the model of log(VOT) for voiceless stops.

Group	Variable	Est. variance	Est. *SD*
Speaker	INTERCEPT	0.018	0.135183
	FOLLOWING VOWEL HEIGHT	0.0016	0.03946
	PLACE OF ARTICULATION$_1$	0.0064	0.079776
	PLACE OF ARTICULATION$_2$	0.00099	0.031413
	PHRASE POSITION	0.0020	0.045087
	SPEAKING RATE DEVIATION	0.0005	0.022357
	FREQUENCY	0.000017	0.004056
Word	INTERCEPT	2.36E-02	0.153764
	PHRASE POSITION	0	0
	MEAN SPEAKING RATE	3.00E-03	0.05479
	DECADE OF BIRTH$_1$	7.64E-03	0.087404
	DECADE OF BIRTH$_2$	8.16E-04	0.028566
	DECADE OF BIRTH$_3$	0	0
	DECADE OF BIRTH$_4$	2.70E-05	0.0052
	DECADE OF BIRTH$_5$	3.96E-04	0.019897

(*continued*)

Table 9: (*continued*)

Group	Variable	Est. variance	Est. *SD*
	SPEAKING RATE DEVIATION	0	0
	DECADE OF BIRTH$_1$:SPEAKING RATE DEVIATION	2.99E-03	0.054651
	DECADE OF BIRTH$_2$:SPEAKING RATE DEVIATION	0	0
	DECADE OF BIRTH$_3$:SPEAKING RATE DEVIATION	0	0
	DECADE OF BIRTH$_4$:SPEAKING RATE DEVIATION	6.34E-05	0.007963
	DECADE OF BIRTH$_5$:SPEAKING RATE DEVIATION	4.23E-05	0.006506

References

Aitken, A. J. 1984. Scots and English in Scotland. In Peter Trudgill (ed.), *Language in the British Isles*, 517–532. Cambridge: Cambridge University Press.

Allen, J. Sean, Joanne L. Miller & David DeSteno. 2003. Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America* 113. 544–552.

Auzou, Pascal, Canan Ozsancak, Richard Morris, Mary Jan, Francis Eustache & Didier Hannequin. 2000. Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics & Phonetics* 14(2). 131–150.

Baayen, R. Harald. 2008. *Analyzing linguistic data*. Cambridge: Cambridge University Press.

Baayen, R. Harald., Richard Piepenbrock & Hedderik van Rijn. 1996. *CELEX2*. Philadelphia: Linguistic Data Consortium.

Balukas, Colleen & Christian Koops. 2014. Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*. doi: 10.1177/1367006913516035

Baran, Jane, Marsha Laufer & 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 & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7.

Belsley, David, Edwin Kuh & Roy Welsch. 1980. *Regression diagnostics: Identifying influential data and sources of collinearity*.Chicago: John Wiley & Sons.

Benjamin, Barbaranne. 1982. Phonological performance in gerontological speech. *Journal of Psycholinguistic Research* 11. 159–167.

Berry, Jeff & Maura Moyle. 2011. Covariation among vowel height effects on acoustic measures. *Journal of the Acoustical Society of America* 130(5). 365–371.

Caramazza, Alfonso, Grace Yeni-Komshian, Edgar B. Zurif & Ettore Carbone. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French–English bilinguals. *Journal of the Acoustical Society of America* 54. 421–428.

Catford, John, C. 2002. *A practical introduction to phonetics*, 2nd edn. Oxford: Oxford University Press.

Cho, Taehong & Peter Ladefoged. 1999. Variations and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27. 207–229.

Cole, Jennifer, Heejin Kim, Hansook Choi & 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(2). 180–209.

Crystal, Thomas & Arthur House. 1988. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America* 83(4). 1553–1573.

Docherty, Gerard. 1992. *The timing of voicing in British English obstruents*. Berlin & New York: Foris.

Docherty, Gerard, Dominic Watt, Carmen Llamas, Damien Hall & Jennifer Nycz. 2011. Variation in voice onset time along the Scottish-English border. In *Proceedings ICPhS XVII*.

Esposito, Anna. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59. 197–231.

Flege, James & Wieke Eefting. 1987. Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15. 67–83.

Fromont, Robert & Jennifer Hay 2012. LaBB-CAT: An annotation store. In *Proceedings of Australasian Language Technology Association Workshop* 10. 113–117.

Geiger, Steven & Joseph Salmons. 2006. Reconstructing variation at shallow time depths: The historical phonetics of 19th-century German dialects in the U.S. In T. Cravens (ed.), *Variation and reconstruction*, 37–58. Amsterdam: Benjamins.

Gelman, Andrew, Jennifer Hill & MasanaoYajima. 2012. Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness* 5(2). 189–211.

Gósy, Maria. 2001. The VOT of the Hungarian voiceless plosives in words and in spontaneous speech. *International Journal of Speech Technology* 4(1). 75–85.

Grant, William 1912. *The pronunciation of English in Scotland*. Cambridge: Cambridge University Press.

Henry, Katherine, Morgan Sonderegger & Joseph Keshet. 2012. Automatic measurement of positive and negative voice onset time. In *Proceedings of INTERSPEECH 2012*.

Hothorn, T., F. Bretz & P. Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3). 346–363.

Iverson, Gregory & Joseph Salmons. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12. 369–96.

Johnston, Paul. 1983. Irregular style variation patterns in Edinburgh speech. *Scottish Language* 2. 1–19.

Johnston, Paul. 1997. Regional variation. In Charles Jones (ed.), *The Edinburgh history of the Scots language*, 433–513. Edinburgh: Edinburgh University Press.

Keshet, Joseph, Morgan Sonderegger & Thea Knowles. 2014. AutoVOT: A tool for automatic measurement of voice onset time using discriminative structured prediction [Computer program]. Version 0.91. http://github.com/mlml/autovot (accessed February 2014).

Kessinger, Rachel & Sheila Blumstein. 1997. Effects of speaking rate on voice onset time in Thai, French, and English. *Journal of Phonetics* 23. 148–68.

Klatt, Dennis. 1975. Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech, Language and Hearing* 18. 686–706.

Krull, Diana. 1991. VOT in spontaneous speech and in citation form words. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm* (*PERILUS*) 12. 101–107.

Kuznetsova, A., P. B. Brockhoff & R. H. B. Christensen. 2014. *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package)*. R package version 2.0–11.

Lenth, Russell V. 2014. *lsmeans*: *Least-Squares Means*. R package version 2.12.

Lisker, Leigh & Arthur Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422.

Lisker, Leigh & Arthur Abramson. 1967. Some effects of context on voice onset time in English stops. *Language and Speech* 10. 1–28.

Masuya, Yoshiro. 1997. Voice onset time of the syllable-initial /p/, /t/ and /k/ followed by an accented vowel in Lowland Scottish English. In *Onseigaku to oninron: shuyo ronko* [Phonetics and phonology: selected papers], 139–172. Tokyo: Kobian Shobo.

Miller, Joanne, Kerry Green & Adam Reeves. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43(1–3). 106–115.

Morris, Richard, Christopher McCrea & Kaileen Herring. 2008. Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics* 36(2). 308–317.

Mortensen, Johannes & John Tøndering. 2013. The effect of vowel height on voice onset time in stop consonants in CV sequences in spontaneous Danish. In R. Eklund (ed.), *Proceedings of Fonetik 2013*, The XXVIth Annual Phonetics Meeting 12–13 June 2013, *Linköping University, Linköping, Sweden: Studies in Language and Culture* 21. 49–520.

Nearey, Terrance & Bernard Rochet. 1994. Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association* 24. 1–18.

Neiman, Gary, Richard Klich & Elaine Shuey. 1983. Voice onset time in young and 70-year-old women. *Journal of Speech and Hearing Research* 26. 118–123.

Nielsen, Kuniko. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39(2). 132–142.

Petrosino, Linda, Roger Colcord, Karen Kurcz & Robert Yonker. 1993. Voice onset time of velar stop productions in aged speakers. *Perceptual and Motor Skills* 76. 83–88.

Podesva, Robert J., Jermay Jamsu Reynolds, Patrick Callier & Jessica Baptiste. 2015. Constraints on the social meaning of released /t/: A production and perception study of U.S. politicians. *Language Variation and Change* 27(1). 59–87.

Ryalls, John, Marni Simon & Jerry Thomason. 2004. Voice onset time production in older Caucasian- and African-Americans. *Journal of Multilingual Communication Disorders* 2. 61–67.

Ryalls, John, Allison Zipprer & Penelope Baldauff. 1997. A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language, and Hearing Research* 40. 642–645.

Sancier, Michele & Carol Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25(4). 421–436.

Sankoff, Gillian & Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83. 560–588.

Scobbie, James M. 2006. Flexibility in the face of incompatible English VOT systems. In Louis M. Goldstein, Doug H. Whalen & Catherine Best (eds.), *Laboratory phonology 8*, 367–392. Berlin: Mouton de Gruyter.

Shockley, Kevin, Laura Sabadini & Carol Fowler. 2004. Imitation in shadowing words. *Perception & Psychophysics* 66(3). 422–429.

Sonderegger, Morgan. 2012. Phonetic and phonological dynamics on reality television. Chicago, IL: University of Chicago Ph.D. thesis.
Sonderegger, Morgan. 2015. Trajectories of voice onset time in spontaneous speech on reality TV. In The Scottish Consortium for ICPhS (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*.
Sonderegger, Morgan & Joseph Keshet. 2012. Automatic measurement of voice onset time using discriminative structured prediction. *Journal of the Acoustical Society of America* 132. 3965–3979.
Stuart-Smith, Jane. 2003. The phonology of modern Urban Scots. In John Corbett, J. Derrick McClure & Jane Stuart-Smith (eds.), *The Edinburgh companion to Scots*, 110–137. Edinburgh: Edinburgh University Press.
Stuart-Smith, Jane. 2007. Empirical evidence for gendered speech production: /s/ in Glaswegian. In Jennifer Cole & Jose Hualde (eds.), *Change in phonology (laboratory phonology 9)*, 65–86. Berlin: Mouton de Gruyter.
Stuart-Smith, Jane, Tamara Rathcke, Morgan Sonderegger & Rachel Macdonald. 2015. A real-time study of plosives in Glaswegian using an automatic measurement algorithm: Change or age-grading? In Eivind Torgersen, S. Hårstad, B. Mæhlum & U. Røyneland (eds.), *Language Variation – European Perspectives V. Selected Papers from the 7th International Conference on Language Variation in Europe (ICLaVE 7), Trondheim, June 2013*, 225–237. Amsterdam/New York: John Benjamins.
Stuart-Smith, Jane, Claire Timmins, Gwilym Pryce & Barrie Gunter. 2013. Television is also a factor in language change: Evidence from an urban dialect. *Language* 89(3). 1–36.
Stuart-Smith, Jane, Claire Timmins & Fiona Tweedie. 2006. Conservation and innovation in a traditional dialect: L-vocalization in Glaswegian. *English World Wide* 27(1). 71–87.
Stuart-Smith, Jane, Claire Timmins & Fiona Tweedie. 2007. Talkin' Jockney: Accent change in Glaswegian. *Journal of Sociolinguistics* 11. 221–261.
Summerfield, Quentin. 1975. Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables. *Speech Perception* 2(4). 61–72. Belfast: Department of Psychology, Queen's University of Belfast.
Takada, Mieko. 2012. VOT variations in Japanese word-initial stops. *Papers from the First International Conference on Asian Geolinguistics*, 271–282. http://agsj.jimdo.com/picag-1/ (accessed September 2014).
Theodore, Rachel, Joanne Miller & David DeSteno. 2009. Individual talker differences in voice-onset-time: Contextual influences. *Journal of the Acoustical Society of America* 125(6). 3974–3982.
Torre, Peter & Jessica Barlow. 2009. Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders* 42(5). 324–333.
Tucker, Ben. 2007. Spoken word recognition of the reduced American English flap. Tucson, AZ: University of Arizona Ph.D. thesis.
van Heuven, Walter, Pawel Mandera, Emmanuel Keuleers & Marc Brysbaert. 2014. Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology* 67. 1176–1190.
Yao, Yao. 2009. Understanding VOT variation in spontaneous speech. In M. Pak (ed.), *Current numbers in unity and diversity of languages*, 1122–1137. Seoul: Linguistic Society of Korea
Yu, Alan C. L., Carissa Abrego-Collier & Morgan Sonderegger. 2013. Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLOS ONE* 8(9). e74746.
Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.