CONSTRAINTS ON THE VARIABILITY IN ENGLISH VOICE ONSET TIME PRODUCTION: EVIDENCE FROM SPONTANEOUS SPEECH

Nguyen Thi Quyen*

School of Languages, International University, Vietnam National University Ho Chi Minh City, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam

Received 15 May 2023
Revised 7 August 2023; Accepted 20 August 2023

Abstract: Voice onset time (VOT) - an aspect of stop production, is known to be constrained by a number of word-level and speaker-level factors. While there has been extensive experimental research on VOT production, few studies have situated this aspect in naturally-occurring spontaneous speech. The goal of this study is twofold. First, we determine if the factors that affect VOT production in experimental studies conducted in speech laboratories are also relevant for VOT production in spontaneous speech. Second, we explore the possible interactions among those factors. In this study, a spoken corpus consisting of clips from a reality TV show was analyzed using a semi-automatic VOT measurement method that allows for the quick and reliable processing of large numbers of VOT measures. The findings confirm the effects of word-level constraints in the expected directions; however, we find negligible effects of speakers' individual differences, namely speech rate and speaker's gender, on VOTs. The current study sheds new light on a range of factors that were previously identified as significantly affecting the patterning of VOTs found in experimentally-elicited data.

Keywords: VOT, stop production, spontaneous speech

1. Introduction

The variability of sound production is conditioned by various phonetic and phonological factors (e.g., phonological context, the language-specific realization of the contrasts, and coarticulatory constraints) and speakers' individual differences. One way to examine this variability is by analyzing phonetic cues, such as voice onset time (VOT). VOT measures the timing between the release of the stop closure and the onset of voicing in the following sonorous segments (e.g., vowels, glides and liquids), and is commonly used to distinguish between voiced stops on the one hand and aspirated and unaspirated stops on the other in many languages, including English. In English, the phonetic realization of the /p/ sound is typically more aspirated than that of the /b/ sound, which is usually phonetically devoiced, resulting in a longer positive VOT value for [p] compared to [b]. These English voiceless stops generally exhibit VOT durations ranging from 40 ms to 100 ms. Meanwhile, voiced stops in English are often devoiced with the exception of those occurring in intervocalic positions, resulting in negative VOTs since the vocal cords start vibrating before the release of the stops (Forrest et al., 1989; Klatt, 1975; Lisker & Abramson, 1964).

A host of factors are known to influence VOT measures, including linguistic factors

* Corresponding author.
Email address: ntquyen@hcmiu.edu.vn

such as place of articulation of the stop sounds, the position of the stop sounds in a syllable, stress placement (Cho & Ladefoged, 1999; Klatt, 1975; Whiteside et al., 2004), lexical frequency and non-linguistic factors such as speaker gender, individual speaking rate, age and emotional state (Koenig, 2000; Ryalls et al., 2004). The location in the vocal tract where the stop consonant is produced has been shown to affect VOT, with longer VOTs measured for velars than for coronals and labials (Lisker & Abramson, 1967; Volaitis & Miller, 1992). The context in which the stop consonant appears also affects VOT, with voiceless stops in initial position having longer VOT values than those in medial or final position. Speaker characteristics such as age, gender, and dialect can also influence VOT, with males having longer VOTs than females and older speakers having shorter VOTs than younger ones. The rate of speech, voice quality, and emotional state of the speaker also affect VOT, with faster speech and creaky voice resulting in shorter VOT values, while anxious or excited speakers tend to have shorter VOTs than calm or relaxed ones.

Previous research has primarily examined the effects of these factors on VOT using carefully controlled speech elicited in phonetic labs, such as read word lists, read sentences, or read passages. However, how these factors influence VOT in naturally occurring spontaneous speech remains unclear, largely due to the difficulty in deriving phonetically robust VOT measures from conversational speech, which can be more time-consuming. Thus, the aim of this study is to investigate if the aforementioned factors affect VOT in spontaneous speech similarly to controlled experiments and to identify any potential interactions among these factors. Studying the variability of VOT production in spontaneous speech is important as it can shed light on how speakers produce speech in natural settings and help gain a better understanding of the complex nature of speech production and its variability.

2. Constraints on the Variability in VOT Production

2.1. Linguistic Factors

Variation in VOT is constrained by a complicated set of linguistic factors (Auzou et al., 2000, Docherty, 1992). Phonological voicing and aspirates are the most significant factors affecting VOT, with VOT behaving differently for voiced and voiceless stops (voiced < unaspirated voiceless < aspirated voiceless). Furthermore, place of articulation heavily influences VOT durations, with VOT expected to increase progressively for bilabial, alveolar, and velar stops (Docherty, 1992; Lisker & Abramson, 1967; Nearey & Rochet, 1994). The phonological context also affects VOT, including following vowel height and following segment identity. Several experimental studies conducted in laboratories have observed that VOT durations are longer before high vowels than low vowels (Berry & Moyle, 2011; Klatt, 1975). Additionally, VOT durations are also found to be longer before consonants in complex onsets than before vowels in CV syllables (Docherty, 1992; Nearey & Rochet, 1994). These findings suggest that the phonetic context in which stop consonants are produced can impact VOT durations.

Recently, there have been investigations into various aspects of words that could affect VOT durations, but some findings remain inconclusive. In one study, Yao (2009) analyzed unplanned American spontaneous speech and discovered that more frequently used words have shorter VOT durations. However, in a replication study that employed read word lists, Yu et al. (2013) found no significant impact of lexical frequency on VOT. Another factor that could be relevant is syllable stress, which has been observed to lead to longer VOT durations, as noted by Cole et al. (2007) and Stuart-Smith et al. (2015). However, this trend is more consistent for

voiceless stops than for voiced counterparts. Similarly, speech rate has a voiced-voiceless asymmetry, with VOT decreasing significantly as speech rate increases for voiceless stops, but not for voiced stops, as observed in studies by Miller et al. (1986) and Stuart-Smith et al. (2015).

2.2. Speaker-related Factors

VOT can be influenced by various speaker-related factors such as age, ethnicity, and gender, and their interactions can also play a role. For example, Ryalls et al. (1997) and Ryalls et al. (2004) examined the effect of ethnicity and gender in younger and older African-American and Caucasian-American male and female speakers and found significant differences in VOT among younger speakers, with male and African-American speakers producing more voicing for voiced stops. However, in older speakers, no significant effects of ethnicity or gender were observed. Turning to the age factor, although several studies have investigated the relationship between VOT and aging, the results have been inconclusive. While Ryalls et al. (2004) reported that older speakers (over 70) have shorter VOT durations than younger speakers, other studies, such as Petrosino et al. (1993) and Torre and Barlow (2009), have found no significant age-related differences in VOT or complex interactions between age and gender. It follows that VOT values may reflect age both as a socially-conditioned life stage and as a physiological consequence of aging. Table 1 below shows a summary of factors constraining VOT durations.

Table 1

Factors Constraining VOT Durations

anterior	posterior
low vowels	high vowels
preceding vowel	preceding consonant
unstressed syllables	stressed syllables
high lexical frequency	low lexical frequency
high speech rate	low speech rate
male	female

3. The Study

The above discussion on the different constraints on VOT measures in experimental and carefully controlled studies leads us to the question whether these same constraints affect VOT production in spontaneous speech and how they interact with each other. In this paper, we aim to investigate the extent to which these constraints affect the duration of VOT in word-initial voiceless stops only. The reason for including only word-initial voiceless stops is that initial voiced stops are frequently devoiced in English but not always, and closure voicing is still observed in their phonetic realization, which can cause complications for annotation and analysis. By focusing on this specific aspect, we hope to gain a deeper understanding of the influence of factors such as lexical frequency, syllable stress, and speaker-related factors like ethnicity, age, and gender on VOT durations. Through this investigation, we hope to provide new insights into the mechanisms that underlie VOT production in spontaneous speech and the various factors that shape it.

3.1. Sample Description

For the purposes of examining natural VOT production, speech data was collected from seasons 30-35 (2020) of the reality TV show The Challenge. The contestants on this show play in teams to compete against each other in a variety of missions to win prizes and advance in the game. Around 25 minutes of audio was extracted from the show, featuring twenty different contestants who speak different dialects of American English. Demographic information such as age, gender, and dialect spoken is provided in Table 2, along with the amount of data collected from each contestant. The exchanges between speakers are brief and consist mainly of questions and answers, with the overall speech register being casual and conversational. This type of spontaneous spoken data offers a valuable opportunity to investigate VOT production in the most natural context possible.

Table 2

Demographic Information of Speakers and Amount of Data for Each Speaker

	Speakers	**Gender**	**Age**	**Dialect spoken**	**Total length of speaker's initial voiceless stops**
1	P.P	M	20	North Central English	0:59
2	G.H	M	23	Inland Northern English	1:33
3	S.B	M	23	Metropolitan New York	1:34
4	V.P	M	20	Metropolitan New York	0:35
5	S.B	M	23	Metropolitan New York	1:56
6	L.R	M	21	Northeastern English	1:07
7	J.R	M	27	Pennsylvania	1:05
8	L.J	M	25	Philadelphia English	0:55
9	K.H	M	25	Philadelphia English	0:48
10	G.H	M	22	Southern	1:27
11	L.P	M	39	New Orleans	0:52
12	C.A	M	25	Western American English	1:21
13	N.A	F	26	Metropolitan New York	1:42
14	M.A	F	24	Northeastern English	1:10
15	N.D	F	38	Northeastern English	1:17
16	D.N	F	31	Northeastern English	1:12
17	D.W	F	32	Northeastern English	1:01
18	A.C	F	25	North Central English	1:51
19	K.J	F	26	Philadelphia English	0:57
20	C.V	F	35	Western American English	1:32
				Total	**24:54**

3.2. Data Preprocessing

Before force-aligning each audio clip from The Challenge seasons 30-35 (2020), two research assistants cross-checked both the orthographic transcription and audio. Afterward, a version of the HTK-based aligner from FAVE (Rosenfelder et al., 2011) was utilized for the force-alignment process. VOT measurements (in ms) for all word-initial voiceless stops (n = 18,437) were semi-automatically conducted following a two-step procedure outlined in Stuart-Smith et al. (2015). AutoVOT software (Keshet et al., 2014) was used to measure VOTs for voiceless stops in the first step, and the results were manually inspected, corrected, and coded by three other research assistants in the second step.

The coding scheme consisting of three labels was utilized by three annotators. The first label indicated that the automatic prediction was correct, while the second label indicated that the automatic prediction was incorrect but could be easily corrected, and therefore was corrected. The third label was used when the data was unusable due to errors in alignment, speaker overlap, background noise, or transcription errors. Inter-coder agreement was determined by manually correcting 10% of the VOT data that was automatically predicted. Predictions from AutoVOT were corrected for 2,121 voiceless stops, which accounts for approximately 12.2% of the total predicted VOT measurements. Unusable VOT measurements for all three voiceless sounds accounted for 29.6% of the total data. Ultimately, the final dataset consisted of 12,272 tokens, including 2,786 [p] tokens, 5,379 [t] tokens, and 4,107 [k] tokens, after excluding the unusable VOT measurements. Table 3 illustrates the breakdown of tokens by the three labels.

Table 3

Number and Proportion of Voiceless Stops Automatically Measured

Voiceless stops	n	Correct	Corrected	Not usable	Final datasets (n)
[p]	2,994	2,063 (68.9%)	723 (24.1%)	208 (7.0%)	2,786
[t]	9,313	4,627 (49.7%)	752 (8.1%)	3,934 (42.2%)	5,379
[k]	5,130	3,461(67.5%)	646 (12.6%)	1,023 (19.9%)	4,107
Total	17,437	10,151 (58.2%)	2,121 (12.2%)	5,165 (29.6%)	12,272

The two variables in our study, speaking rate and lexical frequency, were operationalized as follows. The rate at which a person speaks was determined by dividing the number of syllables in a phrase, which was identified as a portion of speech separated by at least 60 milliseconds of silence or non-speech, by the duration of that phrase in seconds. Kendall's (2013) approach was adopted to exclude pauses from the speaking rate calculation, with a cutoff of 60 ms. As for lexical frequency, it was determined by referring to Subtlex-UK, a new and improved word frequency database for British English, and transforming the frequency values into logarithmic scale after retrieving the orthographic form for each token.

3.3. Statistical Analysis

The aim of this study is to investigate the correlations between VOT duration and seven variables outlined in Section 2. We utilized mixed-effects linear regression models with the lme4 package (Bates et al., 2014) in R (R Core Team, 2014) to achieve this, where VOT was represented as a function of the variables that were discussed earlier. Because we were mainly

interested in voiceless stops with positive VOT values, which could result in an imbalanced distribution of VOT, we used the logarithm of VOT as the dependent variable in the models. The fixed and random effects incorporated into the models are described in detail below.

3.3.1. Fixed Effects

The seven main effects, including five word-level variables and two speaker-level variables were included in the models. Specifications are detailed in Table 4 below:

Table 4

Type and Levels of Variables Included in the Models

	Variables	**Type (levels)**
Word-level	Place of articulation	factor (bilabial, alveolar, velar)
	Following vowel	factor (non-high, high)
	Following segment identity	factor (vowel, consonant)
	Syllable stress	factor (unstressed, stressed)
	Lexical frequency	continuous
Speaker-level	Speaking rate mean	continuous
	Gender	factor (male, female)

Helmert contrasts were used to code categorical variables with the intention of reducing collinearity and making it easier to understand the main effect terms in the models. Table 3 provides information about the levels of each variable used. In order to answer the research questions, we included main effect terms for the seven variables in our models. The continuous variables, Lexical frequency and Speaking rate mean, were centered by subtracting their mean. We evaluated the possibility of interactions between the variables by (1) generating graphs to identify potential interactions in cases where one variable appeared to adjust the other's effect on VOT, and (2) conducting stepwise backward model selection using the step() function in the lmerTest package in R to explore all potential two-way and three-way interactions between the seven variables.

3.3.2. Random Effects

To account for the non-independence of tokens from individual words (12,272 tokens vs. 950 types), random effects were incorporated in the models for *words* and *speakers*. Random intercepts for both *words* and *speakers* were included, as recommended by Allen et al. (2003) and Sonderegger (2012), to address variations among speakers and words, along with other sources of variability. In each model, every possible by-word and by-speaker random slope was included to accommodate the variation among speakers and words in the impacts on VOT that fixed-effect terms captured. This method also prevents Type I error in the fixed-effect coefficients. Furthermore, by including both speaker random intercepts and slopes, it provides some control over factors that were not considered fixed factors in the models.

4. Results

4.1. Exploratory data analysis

The relationship between VOT and the seven variables mentioned above is shown in Figures 1-7. Figure 1 shows plots of the relationship between VOT and gender. From this figure it can be seen that female speakers generally have longer VOT although it appears that male speakers vary in terms of VOT duration. Figure 2 shows plots of the relationship between speaking rate and VOT. There seems to be a small negative relationship between VOT and speaking rate as indicated by the relatively flat trend line as well as big standard error of the trend line.

Figure 1

Relationship Between VOT and Gender

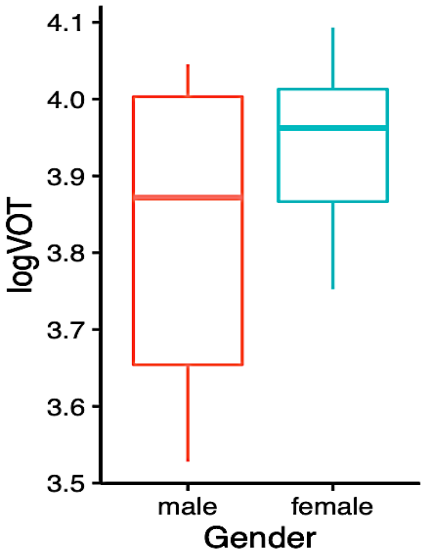


Figure 2

Relationship Between VOT and Speaking Rate

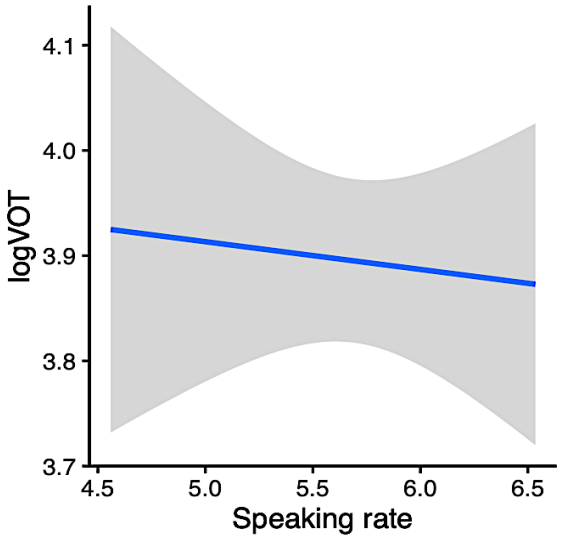


Figure 3

Place of Articulation and VOT

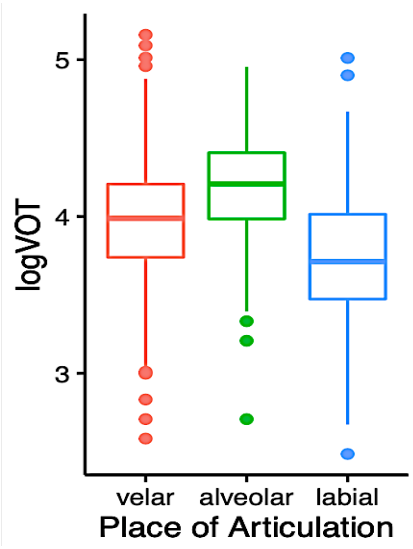


Figure 4

Vowel Height and VOT

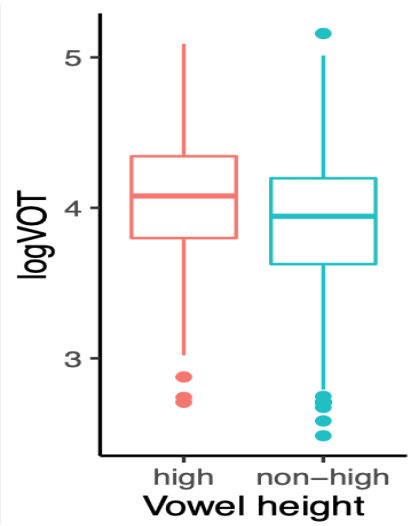


Figure 3 shows plots of the relationship between place of articulation and VOT production. As can be seen from Figure 3, the place of articulation of a stop can have an effect on its VOT production. Velar and alveolar stops appear to be produced with longer VOT compared to bilabial stops. However, highest VOT durations are observed with alveolar stops, which is unexpected given the earlier discussion. Figure 4 displays the relationship between VOT and vowel height in which VOT is larger with high vowels and smaller with low vowels, which is as expected.

Figure 5

Following Segment Identity and VOT

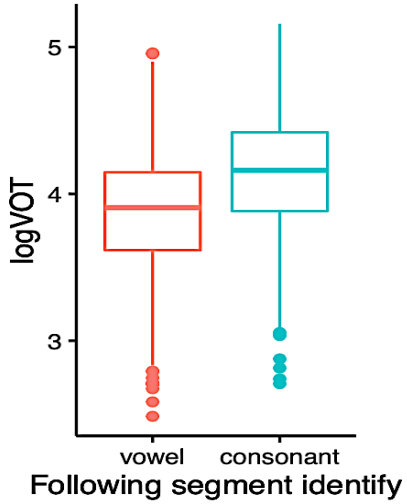


Figure 6

Syllable Stress and VOT

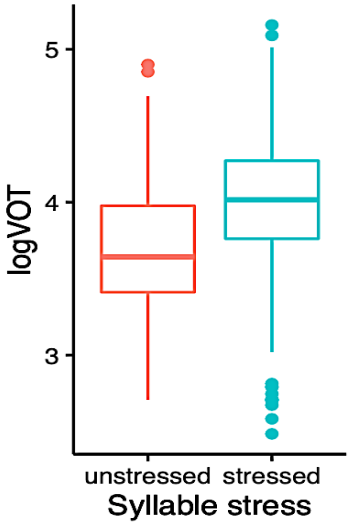
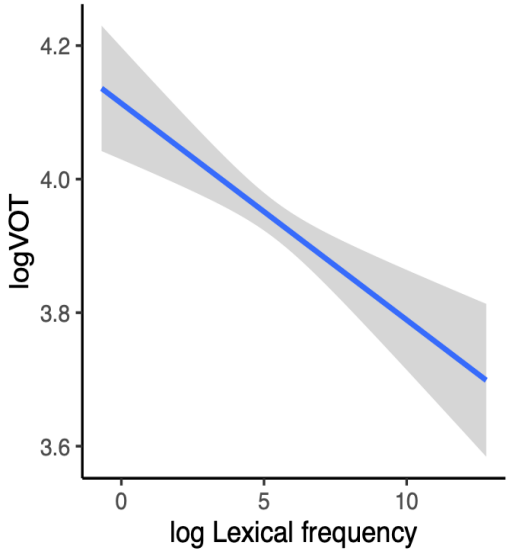


Figure 7

Lexical Frequency and VOT



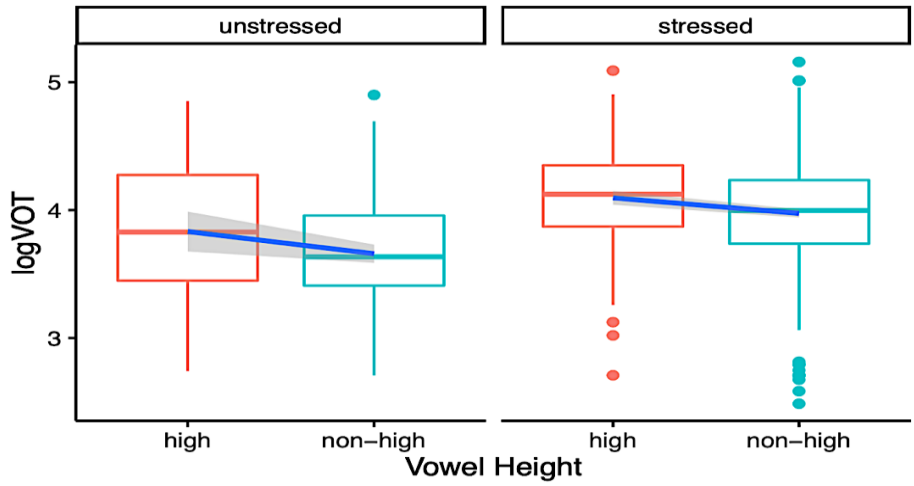
Another word-level factor, following segment identity, appears to modulate VOT as well, as can be seen in Figure 5. A stop consonant is likely to be produced with a longer VOT when it is followed by a consonant (in a consonant cluster) than by a vowel (in a CV syllable). Syllable stress appears to modulate VOT duration as well, as can be seen in Figure 6. It can be seen that when a stop consonant occurs in a stressed syllable it will be produced with a larger VOT. Lastly, as can be seen from Figure 7, larger VOT is observed in words with lower

frequency, indicating an inverse relationship between VOT and word frequency.

To explore possible interactions among the variables, we plot all two-way and three-way interactions, showing the trend lines for mean VOTs in each interaction. For reasons of space, we only show figures which contain possible interactions, manifested by the divergent trend lines for mean VOTs. In Figure 8, the two trend lines are quite divergent, meaning the effect of vowel height on VOT varies depending on whether that the stop occurs in a stressed syllable or in an unstressed syllable. Such observation suggests a strong interaction between vowel height and syllable stress.

Figure 8

Interaction Between Following Vowel Height and Syllable Stress



Similar observations can be made from Figure 9 below, which indicates a possible interaction between vowel height and word frequency. The effect of lexical frequency on VOT duration appears to be modulated by the height of vowel following the stop. Another interaction can be observed between following segment identity and word frequency, as shown in Figure 10. Likewise, the effect of lexical frequency on VOT varies depending on the identity of the segment following the stop.

Figure 9

Interaction Between Vowel Height and Lexical Frequency

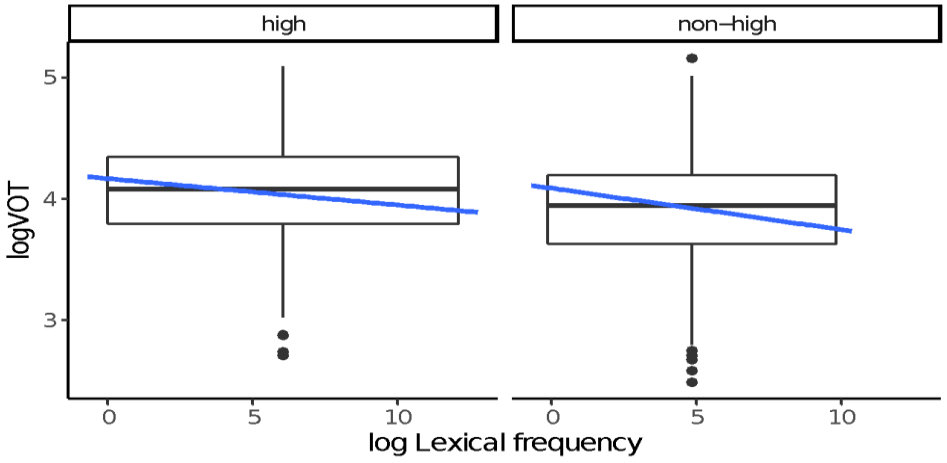
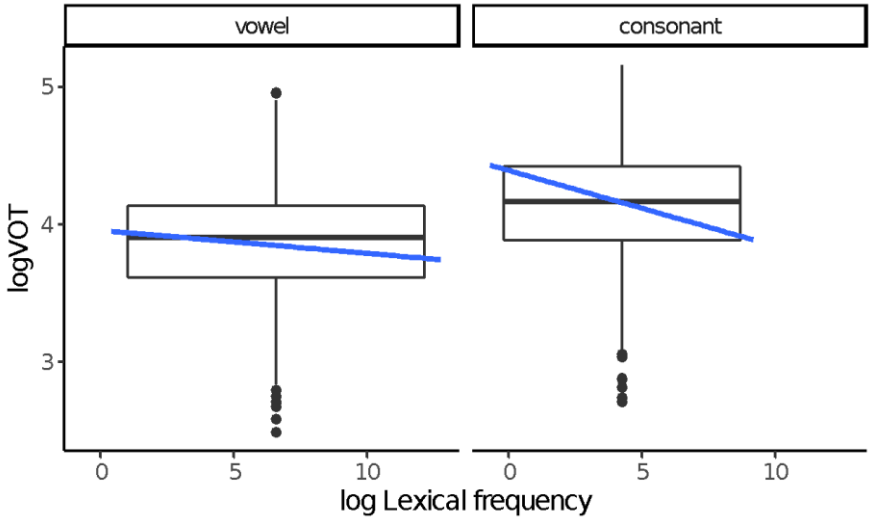


Figure 10

Interaction Between Following Segment Identity and Lexical Frequency



4.2. Results From the Mixed-Effects Linear Model

Based on the above exploratory analysis, a mixed-effects linear regression model of VOT will include seven fixed-effects variables mentioned earlier and three interactions between vowel height and syllable stress, vowel height and lexical frequency, and following segment identity and lexical frequency. Also, in order for the coefficients of all effects to be comparable, all the values of variables are standardized before fitting into the model. The two by-word and by-subject random terms do not significantly affect VOT[1]. In Table 5 we only report the results for fixed-effect terms.

Table 5

Mixed-effects Model Summary

Coefficient	EST	SE	*df*	*t*	*p*-value
(intercept)	3.903	0.042	33.4	92.365	0.001***
Gender	0.109	0.071	20.4	1.551	0.136
Speaking rate	-0.002	0.073	21	-0.033	0.974
Vowel height	-0.131	0.042	423	-3.105	0.002***
Place of articulation	-0.177	0.018	756.4	-9.794	0.001***
Syllable stress	0.377	0.035	1388	10.641	0.001***

[1] We report random effects here, as suggested by an anonymous reviewer.

Random effects:

Groups	Name	Variance	Std. Dev
word	(Intercept)	0.04225	0.2055
speaker	(Intercept)	0.02334	0.1528
Residual		0.16299	0.4037

Following segment identity	0.110	0.043	516	2.556	0.010*
Lexical frequency	-0.123	0.033	609.4	-3.770	0.001***
Vowel height : Syllable stress	-0.208	0.080	1424	-2.597	0.009**
Vowel height : Lexical frequency	-0.138	0.066	63.3	-2.099	0.036*
Lexical frequency : Following segment identity	-0.179	0.068	852.5	-1.606	0.009**

In our analysis of the factors that influence VOT, we initially examined speaker-level variables, such as *gender* and *mean speaking rate*, and found that they did not significantly impact VOT ($p = 0.136$ and $p = 0.974$, respectively), as presented in Table 5. We then proceeded to investigate word-level variables and discovered that all five variables we studied had a significant effect on VOT. Specifically, place of articulation had a significant influence on VOT production ($p = 0.001$). To further explore this relationship, we conducted post-hoc Tukey tests and observed that bilabials had a significantly shorter VOT than alveolars and velars ($p = 0.0001$), while alveolars and velars were not significantly different from each other ($p = 0.39$). Additionally, following vowel heights, following segment identity, word frequency and syllable stress all had statistically significant effects and in the expected direction. Specifically, VOT measures were significantly longer when the word-initial voiceless stops occurred in a stressed syllable of a frequently occuring word, were followed by a high vowel or a consonant.

Our analysis now turns to examining the interaction effects among the variables. Our model predicts a significant interaction between syllable stress and vowel height, indicating that the difference in VOT production between stressed and unstressed syllables is significantly influenced by whether the following vowel is high or non-high ($p = 0.009$). Additionally, we observed a significant interaction effect between vowel height and lexical frequency, where the difference in VOT between more and less frequent words depends on whether the following vowel is high or non-high ($p = 0.002$). Furthermore, the difference in VOT when followed by a consonant versus a vowel is significantly influenced by word frequency ($p = 0.009$). Overall, our mixed-effects linear regression model confirmed the findings from our exploratory data analysis.

5. Discussion

The aim of the current study was to explore constraints on VOT variability in spontaneous speech, which has been relatively understudied compared to VOT in laboratory speech. This is due to the difficulties and time-consuming nature of analyzing VOT in spontaneous data. To overcome this challenge, we adopted a quick and feasible VOT measurement method developed by Stuart-Smith et al. (2015) that employs a semi-automatic procedure with the AutoVOT algorithm to analyze a large number of reliable VOT measures. We also examined both speaker-level and word-level factors that may influence VOT measures. In the following sections, we will compare our findings with previous research on VOT in laboratory speech to gain a better understanding of the extent to which our results align with established findings.

Several laboratory studies have demonstrated that speaking rate has an impact on VOT duration (Kessinger & Blumstein, 1997; Miller et al., 1986). However, our analysis of spontaneous speech data suggests that such variables have an insignificant effect. With respect to speaker gender, previous studies have shown that male speakers tend to produce significantly

lower VOT than female speakers (Koenig, 2000; Morris et al., 2008; Oh, 2021; Ryalls et al., 1997; Whiteside & Irving, 1998; Whiteside & Marshall, 2001), although the results vary depending on age and ethnicity. In our study, we observe a weak relationship between VOT and gender. This may be attributed to our relatively small sample size of only 20, which could make it difficult to detect any trend.

Let's now shift our focus to the word-level factors that impact VOT production. Previous studies conducted under strictly controlled conditions in the laboratory have suggested that VOT variation is limited by the place of articulation. The literature often reports the VOT hierarchy, which suggests that the further back the place of articulation the longer the VOT (Cho & Ladefoged, 1999; Lisker & Abramson, 1964). This hierarchy is generally more noticeable with voiced stops in lab studies. Our analysis of spontaneous speech data revealed that bilabials had shorter VOTs than alveolars and velars, with no significant difference between the latter two, which supports the findings of Docherty (1992). The other word-level factors examined in our study showed significant effects in line with those reported in lab studies. Specifically, VOT was longer before high vowels (Berry & Moyle, 2011; Klatt, 1975) and shorter in more frequent words (Sonderegger, 2012; Yao, 2009). Moreover, syllable stress had a significant effect on VOT, with stops in stressed syllables having a significantly longer VOT than those in unstressed syllables (Cole et al., 2007; Lisker & Abramson, 1967; Stuart-Smith et al., 2015). Lastly, we found that the following segment after a word-initial stop, whether a consonant or a vowel, significantly influenced VOT duration, with VOT being longer when followed by a consonant than by a vowel. This aspect has not been extensively studied in previous research except for Klatt (1975), and thus our findings offer some novel insights that can be corroborated in lab studies.

6. Conclusions

The exploration of VOT in stop production has been comprehensive, although investigations of this aspect in naturally-occurring speech have been limited, owing to the difficulty of measuring VOT in conversational discourse. Nevertheless, we were able to surmount this challenge by implementing a rapid and reliable approach developed by Stuart-Smith et al. (2015), which utilized a semi-automatic process to analyze extensive VOT measurements using the AutoVOT algorithm. Our study demonstrated that word-level constraints had the anticipated impacts on VOT, while we did not detect any significant effects for speaker-level variables, such as speech rate and gender. Overall, our inquiry provides a fresh outlook on the range of factors that restrict VOT patterns in spontaneous speech, and future studies incorporating larger sample sizes or different populations are necessary to scrutinize these effects in greater detail.

References

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice onset time. *The Journal of the Acoustic Society of America, 113*, 544-552. https://doi.org/10.1121/1.1528172

Auzou, P., Ozsancak, C., Morris, R., Jane, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics & Phonetics, 14*(2), 131-150. https://doi.org/10.1080/026992000298878

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using 'Eigen' and S4*. R package version 1.1-7. https://cran.r-project.org/web/packages/lme4/index.html.

Berry, J., & Moyle, M. (2011). Covariation among vowel height effects on acoustic measures. *Journal of the Acoustical Society of America, 130*(5), 365-371. https://doi.org/10.1121%2F1.3651095

Cho, T., & Ladefoged, P. (1999). Variations and universals in VOT: Evidence from 18 languages. *Journal of Phonetics, 27,* 207-229. https://doi.org/10.1006/jpho.1999.0094

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (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.

Docherty, G. (1992). *The timing of voicing in British English obstruents.* Foris. https://doi.org/10.1515/9783110872637

Forrest, K., Weismer, G., & Turner, S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. *Journal of the Acoustical Society of America, 85,* 2608-2622. https://doi.org/10.1121/1.397755

Kendall, T. (2013). *Speech rate, pause and sociolinguistic variation*, Palgrave Macmillan.

Keshet, J., Sonderegger, M., & Knowles, T. (2014). *AutoVOT*: *A tool for automatic measurement of voice onset time using discriminative structured prediction [Computer program]*,Version 0.91. https://doi.org/10.1121/1.4763995

Kessinger, R., & Blumstein, S. (1997). Effects of speaking rate on voice onset time in Thai, French, and English. *Journal of Phonetics, 23,* 148-68. https://psycnet.apa.org/doi/10.1006/jpho.1996.0039

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research, 18,* 686-706. https://doi.org/10.1044/jshr.1804.686

Koenig, L. (2000). Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research, 43*, 1211-1228,

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20,* 384-422.

Lisker, L., & Abramson, A. (1967). Some effects of context on voice onset time in English stops. *Language and Speech, 10,* 1-28. https://doi.org/10.1080/00437956.1964.11659830

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

Morris, R., McCrea, C., & Herring, K. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics, 36*(2), 308-317. http://dx.doi.org/10.1016/j.wocn.2007.06.003

Nearey, T., & Rochet, B. (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. https://doi.org/10.1017/S0025100300004965

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics, 39*(1), 59-67. https://doi.org/10.1016/j.wocn.2010.11.002

Petrosino, L., Colcord, R., Kurcz, K., & Yonker, R. (1993). Voice onset time of velar stop productions in aged speakers. *Perceptual and Motor Skills, 76,* 83-88. https://doi.org/10.2466/pms.1993.76.1.83

R Core Team (2014). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (2011). *FAVE (Forced Alignment and Vowel Extraction).*

Ryalls, J., Zipprer, A., & Baldauff, P. (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. https://doi.org/10.1044/jslhr.4003.642

Ryalls, J., Simon, M., & Thomason, J. (2004). Voice onset time production in older Caucasian and African-Americans. *Journal of Multilingual Communication Disorders, 2,* 61-67. https://doi.org/10.1080/1476967031000090980

Sonderegger, M. (2012). *Phonetic and phonological dynamics on reality television*. University of Chicago dissertation.

Sonderegger, M., & Keshet, J. (2012). Automatic measurement of voice onset time using discriminative structured prediction. *The Journal of the Acoustical Society of America, 132,* 3965-3979. https://doi.org/10.1121/1.4763995

Sonderegger, M. (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*.

Stuart-Smith, J., Sonderegger, M., Rathcke, T., & MacDonald, R. (2015). The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology, 6*, 505-549. https://doi.org/10.1515/lp-2015-0015

Torre, P., & Barlow, J. (2009). Age-related changes in acoustic characteristics of adult speech. *Journal of Communication Disorders, 42*(5), 324-333. https://doi.org/10.1016/j.jcomdis.2009.03.001

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America, 92*(2), 723-735. https://doi.org/10.1121/1.403997

Whiteside, S. P., & Irving, C. J. (1998). Speakers' sex differences in voice onset time: a study of isolated word production. *Perceptual and Motor Skills, 86*(2), 651-654. https://doi.org/10.2466/pms.1997.85.2.459

Whiteside, S. P., & Marshall, J. (2001). Developmental trends in voice onset time: Some evidence for sex differences. *Phonetica, 58*(3), 196-210. https://doi.org/10.1159/000056199

Whiteside, S. P., Henry, L. & Dobbin, R. (2004). Sex differences in voice onset time: a developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America, 116*(2), 1179-1183. https://doi.org/10.1121/1.1768256

Yao, Y. (2009). Understanding VOT variation in spontaneous speech. In M. Pak (Ed.), *Current numbers in unity and diversity of languages* (pp.1122-1137). Linguistic Society of Korea.

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PloS One, 8*(9), e74746. https://doi.org/10.1371/journal.pone.0074746

NHỮNG YẾU TỐ RÀNG BUỘC THỜI GIAN KHỞI THANH CỦA PHỤ ÂM TẮC TRONG TIẾNG ANH: NGHIÊN CỨU TỪ LỜI NÓI TỰ NHIÊN

Nguyễn Thị Quyên

Khoa Ngôn ngữ, Trường Đại học Quốc tế, Đại học Quốc gia TP.HCM,
Khu phố 6, Phường Linh Trung, Thành phố Thủ Đức, Thành phố Hồ Chí Minh, Việt Nam

Tóm tắt: Thời gian khởi thanh (voice onset time - VOT) là một giá trị định lượng độ lệch pha tính bằng mili giây giữa hoạt động đóng/mở của các cơ quan cấu âm trên thanh quản và dây thanh. Như đã biết, đây là một đại lượng quan trọng trong việc phân biệt các phụ âm hữu thanh và phụ âm vô thanh và bật hơi trong nhiều ngôn ngữ trên thế giới nói chung, và tiếng Anh nói riêng. Trong sản sinh phụ âm tắc, có nhiều nghiên cứu cho rằng giá trị đại lượng này bị ràng buộc bởi nhiều khía cạnh về người nói và đặc điểm ngữ âm của môi trường xung quanh phụ âm tắc. Trên thế giới đã có nhiều nghiên cứu về các yếu tố ràng buộc thời gian khởi thanh, nhưng hầu hết các nghiên cứu này đều dựa trên các thí nghiệm được thiết kế nghiêm ngặt trong phòng lab để thu thập ngữ liệu. Có ít nghiên cứu đặt vấn đề này dựa trên lời nói xảy ra một cách tự nhiên mà không thực hiện trong phòng lab. Do đó, nghiên cứu này đặt ra hai mục tiêu như sau: (1) Xác định xem các khía cạnh ảnh hưởng đến thời gian khởi thanh trong các nghiên cứu thực nghiệm trong phòng lab có gây ra ảnh hưởng tương tự trong lời nói tự nhiên hay không; (2) Khám phá những tương tác có thể xảy ra giữa những khía cạnh đó. Trong nghiên cứu này, chúng tôi tập hợp dữ liệu nói gồm các đoạn clip cắt từ một chương trình truyền hình thực tế đã được phân tích bằng một phương pháp đo VOT bán tự động cho phép xử lý nhanh chóng và tin cậy một số lượng lớn các giá trị VOT. Kết quả của nghiên cứu cho thấy các khía cạnh ở cấp độ từ chứa phụ âm vô thanh có ảnh hưởng tới giá trị VOT như đã được xác lập ở các nghiên cứu trong phòng lab. Tuy nhiên, nghiên cứu chỉ ra tác động không đáng kể của sự khác biệt cá nhân của người nói, cụ thể là tốc độ nói và giới tính của người nói, đối với giá trị VOT.

Từ khóa: thời gian khởi thanh, sản sinh phụ âm tắc, lời nói tự nhiên