Understanding VOT Variation in Spontaneous Speech

Yao Yao

Linguistics Department, University of California, Berkeley
1203 Dwinelle Hall #2650, UC Berkeley, CA 94720
yaoyao@berkeley.edu

Abstract

This paper reports a corpus study on the variation of VOT in voiceless stops in spontaneous speech. Two speakers' data from the Buckeye corpus are used: one is an older female speaker with a low speaking rate while the other is a younger male speaker with an extremely high speaking rate. Linear regression analysis shows that place of articulation, word frequency, phonetic context, speech rate and utterance position all have an effect on the length of VOT. However, altogether less than 20% of the variation is explained in both speakers, which suggests that pronunciation variation in spontaneous speech is a highly complicated phenomenon which might need more sophisticated modeling. Our results also show a great deal of individual differences.

Keywords: VOT variation, spontaneous speech, corpus study.

1. Background

Voice onset time (VOT) is the duration between consonant release and the beginning of the vowel. English voiceless stops (i.e. [p], [t], and [k]) typically have VOT durations of 40ms – 100ms (Forrest et al., 1989; Klatt, 1975; Lisker & Abramson, 1964). In the broad literature on English VOT, it has been shown that VOT varies with a number of factors, including linguistic factors (place of articulation, identity the following vowel and speaking rate), and non-linguistic factors (age, gender and other physiological characteristics of the speaker). In this study, we report a corpus study on VOT variation that takes into consideration the features of the target word and the running context. We use two speaker's naturalistic speech data form interviews, and built separate regression models. Our main goal is to study the effect of lexical and contextual factors on VOT in running speech. The comparison of the two models also reveals individual differences between the two speakers.

The most well-studied factor in VOT variation is place of articulation. It has been confirmed in various studies that VOT increases when the point of constriction moves from the lips to the velum, both in isolated word reading and read speech (Zue, 1976; Crystal & House, 1988; Byrd, 1993; among others), and this pattern is not limited to the English language (Cho & Ladefoged, 1999). Speech rate is another conditioning factor. Kessinger and Blumstein (1997, 1998) reported that VOT shortened when speaking rate increases (also see Volaitis & Miller 1992, Allen et al. 2003). It has also been proposed that phonetic context, in particular, the following vowel, has an effect on the length of VOT. Klatt (1975) reported longer VOT before sonorant consonants than before vowels. Klatt also found that voiceless stops typically had longer VOTs when followed by high, close vowels and shorter VOTs when followed by low, open vowels (also see Higgins et al. 1998). In addition, there is also an indirect influence from the following vowel context in that some VOT variation patterns are only observed in certain vowel environments (Neiman et al. 1983; Whiteside et al. 2004).

A different line of research on VOT variation focuses on non-linguistic factors. Whiteside & Irving (1998) studied 36 isolated words spoken by 5 men and 5 women, all in their twenties or thirties, and showed that the female speakers had on average longer VOT than the male speakers. The pattern was confirmed in several other studies (Ryalls et al. 1997; Koenig, 2000; Whiteside & Marshall 2001). Age has also been suggested as a conditioning factor of VOT. Ryalls et al. (1997, 2004) found that older speakers have shorter VOTs than younger speakers, though their syllables have longer durations. A tentative explanation is that older speakers have smaller lung volumes and therefore produce shorter periods of aspiration (see also Hoit et al., 1993). However, no age effect is found in some other studies (Neiman et al., 1983; Petrosino et al., 1993). Other non-linguistic factors that have been studied include ethnic background (Ryalls et al. 1997), dialectal background (Schmidt and Flege, 1996; Syrdal, 1996), presence of speech disorders (Baum & Ryan, 1993; Ryalls et al 1999), and the setting of the experiments (Robb et al., 2005). Last but not least, at least part of the VOT variation is due to idiosyncratic articulatory habits of the speaker. Allen et al's (2003) study shows that after factoring out the effect of speaking rate, the speakers still have different VOTs, though the differences are attenuated.

Despite the large size of the literature on VOT, most of the existing studies use experimental data from single-word productions and therefore typically have a limited set of target syllables and phonetic contexts. (The only two exceptions are Crystal & House [1988] and Byrd [1993], both of which used read speech data from speech corpora.) However, what happens in unplanned spontaneous speech? We know that speakers have more VOT variability in directed conversation than in single-word productions (Lisker & Abramson, 1967; Baran et al., 1977). But does that mean that the conditioning factors are largely the same, only with aggrandized effects or that additional factors are at play? More importantly, what is the general pattern of variation when all factors are present? The current study is a first attempt to address these questions. We use naturalistic data from interviews and build models of VOT variation with features of the word and the running contexts. The features we consider have been suggested in the literature to affect either VOT (such as place of articulation, phonetic context and speaking rate) or pronunciation variation in spontaneous speech in general (such as word frequency and utterance position).

2. Methodology

2.1. Data

The data we use are from the Buckeye Corpus (Pitt et al., 2007), which contains interview recordings from 40 speakers, all local residents of Columbus, OH. Each speaker was interviewed for about an hour with one interviewer. Only the interviewee's speech was digitally recorded in a quiet room with a close-talking head-mounted microphone. At the time of this study, 19 of the 40 speakers' data were available. (In fact, 20 speakers' transcripts were available, but one speaker's data were excluded due to inconsistency in the transcription.) For this study, two of the 19 speakers' data are used. These two speakers, s20 (recoded as F07 in the current study) and s32 (recoded as M08), were selected because they differed from each other in all possible dimensions. F07 is an older female speaker with the lowest speaking rate among all 19 speakers (4.022 syll/s) while M08 is a young male speaker with the highest speaking rate (6.434 syll/s).

Since word-medial stops are often flapped in American English, we limited the dataset to word-initial position only. Speaker F07 has 231 word-initial voiceless stops and speaker M08 has 618 such tokens. An automatic burst detection program was used to find the point of release in each token. More than 57% (N=492) of the tokens were manually checked, and the error was under 3.5ms. 105 tokens (7 of F07 and 98 of M08) were excluded since the automatic program failed to find a reliable point of release in these stop tokens, due to either no closure-release transition or extraordinary multiple releases. (For a detailed discussion on the automatic burst detection program, please see Yao, 2008 in the same volume.) The average VOT of F07 is 57.41ms, with a standard deviation of 26.00ms, while M08's average VOT is 34.86ms, with a standard deviation of 19.82ms. In fact, as shown in Figure 2, M08 has the shorter average VOT of all 19 speakers. The large difference in VOT between the two speakers (~23ms) is probably due to the fact that M08 speaks much faster than F07 in general. Both speakers' VOT values show a great deal of variation (standard deviation > 19ms in both speakers), which will be the focus of the analysis in the rest of the paper.

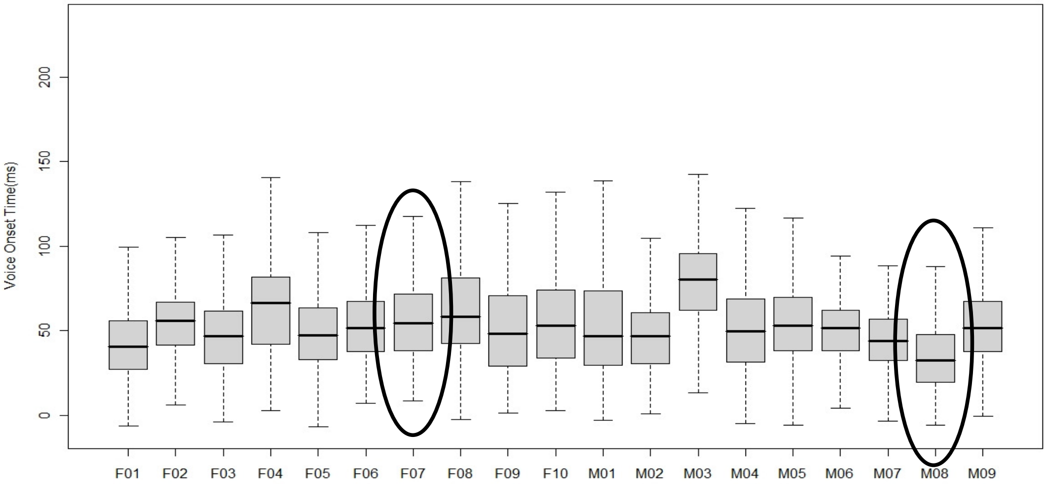

Figure 1. Average VOT of all 19 speakers (F07's and M08's data are circled)

In order to test the effect of surrounding phonetic context, we excluded utterance-initial tokens (14 from F07 and 54 from M08), since the preceding context was not speech sound in these cases. This leaves speaker F07 with 210 tokens and speaker M08 with 466 tokens.

It has been suggested in the literature that content words and function words are processed differently (see Bell et al, to appear and the references in it). In our data, function words have shorter VOTs than content words in both speakers' data (see Figure 2), and the effect still remains after word frequency is controlled for. Since content words comprise the majority of the target tokens (see Table 1), we decided to model the variation of VOT in content words only. Thus, in the final dataset, speaker F07 has 155 tokens and M08 has 346.

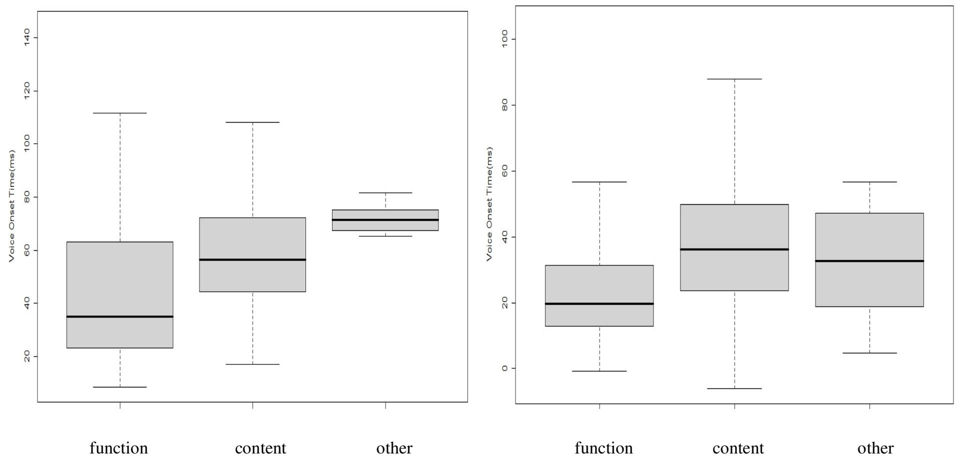

Figure 2. Average VOT by word class in F07 (left) and M08 (right)

	Content	Function	Other
F07	155	47	8
M08	346	104	16

Table 1. Token counts by word class

2.3. Regression model

Linear regression is used to predict the length of VOT in each stop token in the final dataset. Two speakers' data are modeled independently, using the same method. The independent variables that are considered are place of articulation (POA), word frequency, phonetic context, speech rate and utterance position. All predictor variables are added to the model sequentially (in the above order). Adjusted R^2, a model parameter that indicates how much variation is explained, is used to evaluate model performance. The general principle of modeling is that a predictor variable will stay in the model if R^2 is improved significantly. Thus the results that are reported below should be understood as the difference in model performance after adding the current variable, on top of all previously added variables. For some variables, more than one measure is tested and the most significant one is kept in the model.

3. Results

3.1. Effect of POA

The first variable added to the regression model is place of articulation. Various studies have confirmed that VOT in voiceless stops increases as the place of articulation moves backwards, from the lips to the velum. However, this trend is only observed in one of the two speakers in the current study. In speaker F07's data, POA doesn't turn out to be a significant factor for predicting VOT (p=0.216), and doesn't explain any variation at all (R^2=0). Moreover, the average VOT of [p], [t], and [k] doesn't follow the pattern of increasing VOT in more backward stops ([p]=68.56ms; [t]= 61.56ms; [k]= 68.40ms). For speaker M08, on the hand, POA is an important predictor for VOT (p<0.001), with [p] having the shortest VOT (33.14 ms), followed by [t] (44.20 ms) and [k] (47.68 ms), which is consistent with the pattern reported in the literature. POA alone explains about 9.2% of the variation in VOT in M08's data. Figure 3 shows the average VOT of the three stop categories in two speakers. Despite the fact that POA only appears to be a significant predictor in one speaker's data, we decided to keep it in both speakers' models, mostly because it has been claimed to have an important effect on VOT in the literature and it is possible that the effect will show up in the interaction with other predictor variables.

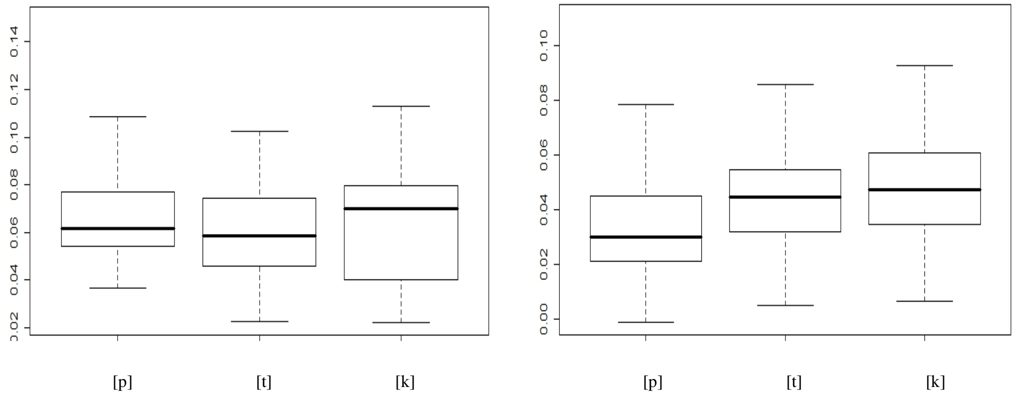

Figure 3. Average VOT by stop category in F07 (left) and M08 (right)

3.2. Effect of word frequency

Word frequency is one of the most well-documented factors on pronunciation variation in spontaneous speech. Frequent words have shorter durations and are more susceptible to various lenition processes, such as vowel reduction, tapping and palatalization, consonant deletion, etc. (Fidelholz, 1975; Fosler-Lussier and Morgan, 1999; Bybee, 2000; Bell et al., 2003, to appear; Pluymaekers et al., 2005; among others).

However, to our knowledge, the effect of word frequency on VOT in connected speech hasn’t been investigated yet.

In the current study, two types of frequency measures are examined: one is the log of the word frequency in the CELEX database (Baayen et al., 1995), the other is the log of the word frequency calculated from the Buckeye corpus over all speakers. Not surprisingly, the two measures are highly correlated (r= 0.826). In both speakers’ data, the Buckeye frequency is a better predictor of VOT than the CELEX frequency (see Table 2). Adding the Buckeye word frequency to the model improves the performance by 1.7% in speaker F07 and 0.4% in speaker M08. In both models, there is a negative relation between word frequency and VOT, i.e. more frequent words have shorter VOTs. But the effect is not very strong, as shown in the relatively small change in R^2.

	Previous R^2 (%)	**Term added**	**New R^2 (%)**
F07	0	Celex frequency	1.3
		Buckeye frequency	1.7
M08	9.2	Celex frequency	9.4
		Buckeye frequency	9.6

Table 2. Change in model performance after adding word frequency

3.3. Effect of phonetic context

For studying the effect of phonetic context, we coded each token for whether the preceding/following phone is a consonant or a vowel. Interestingly, in speaker F07’s data, only the preceding phone contributes to the prediction of VOT (R^2 increases by 0.2%), but not the following one (R^2 decreases by 1.1%); in speaker M08’s data, it is the following phone that predicts VOT (R^2 increases by 0.48%), but not the preceding one (R^2 decreases by 0.33%) (see Table 3). In both speakers’ models, however, the category of the preceding/following phone is not a strong factor in VOT variation (see Figure 4).

	Previous R^2 (%)	**Term added**	**New R^2 (%)**
F07	1.7	Category of the preceding phone	1.9
		Category of the following phone	0.6
M08	9.6	Category of the preceding phone	9.27
		Category of the following phone	10.08

Table 3. Change in model performance after adding phonetic context

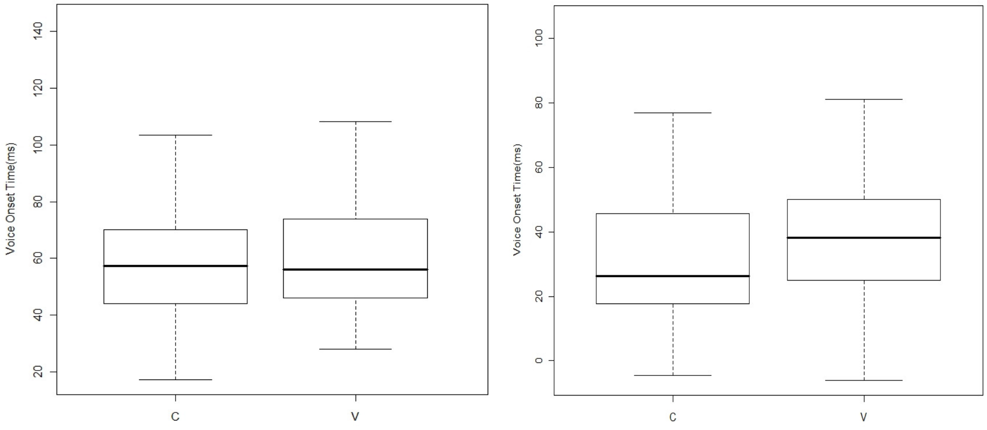

Figure 4. Effect of phonetic context (left: average VOT by category of the preceding phone in F07; right: average VOT by category of the following phone in M08)

3.4. Effect of speech rate

Speech rate is intuitively easy to understand, but hard to measure in practice. Previous studies on the effect of speech rate have been predominantly using the number of syllables produced per second in the local pause-bounded stretch as a measure of the contextual speech rate. In this study, in addition to the stretch speed measure, two more local speed measures are also tested: duration of the next phone (in ms), and the average speed of the three-word chunk centered at the target word (in number of syllables per second).

In both speakers' regression models, all three speed measures improve the performance of the model, predicting that the faster the speech is, the shorter the VOT. However, among the three measures, the average speed of the three-word chunk, is the best VOT predictor in speaker F07's model (R^2 increases by 9.9%), whereas in speaker M08's model, the most local one, i.e. the duration of the next phone, predicts VOT the best (R^2 increases by 6.54%) (see Table 4). Curiously, the most often used speed measure, i.e. average speed of the local stretch, doesn't turn out to be the best speed predictor for VOT in either speaker's model. One might argue that this is at least partly because the calculation of both 3-word-chunk speed and local stretch speed already include the target word (and hence also the predicted VOT) and this inherent correlation is higher in the more local speed measure, which makes it a seemingly better predictor. We agree that a better way to calculate these speed measures is to exclude the target word from the calculation to eliminate the inherent correlation. However, it should also be noted that even with the current calculation, 3-word-chunk speed is not necessarily a better predictor than local stretch speed (as in M08's model), which suggests that despite the different degrees of inherent correlation with VOT, the two speed measures do seem to be measuring different contexts.

	Previous R^2 (%)	Term added	New R^2 (%)
F07	1.9	Duration of the next phone	5.1
		Average speed of the local 3-word chunk	11.8
		Average speed of the local stretch	7.1
M08	10.08	Duration of the next phone	16.62
		Average speed of the local 3-word chunk	12.85
		Average speed of the local stretch	15.07

Table 4. Change in model performance after adding speed measures

3.5. Effect of utterance position

In order to test the effect of utterance position (mostly in the form of utterance final lengthening), each token is coded for whether the immediately following word is silence. Altogether 9 out of the 155 tokens in F07's data and 34 out of 312 in M08's data are coded as followed by pause, i.e. utterance-final. As shown in Figure 5, for speaker F07, utterance-final tokens have significantly longer VOT than utterance-medial tokens, whereas for speaker M08, the two categories have similar average VOTs. Not surprisingly, adding this variable to the regression model improves the performance by 7.31% in F07's model, but decreases it by 3.13% in M08's model (see Table 5).

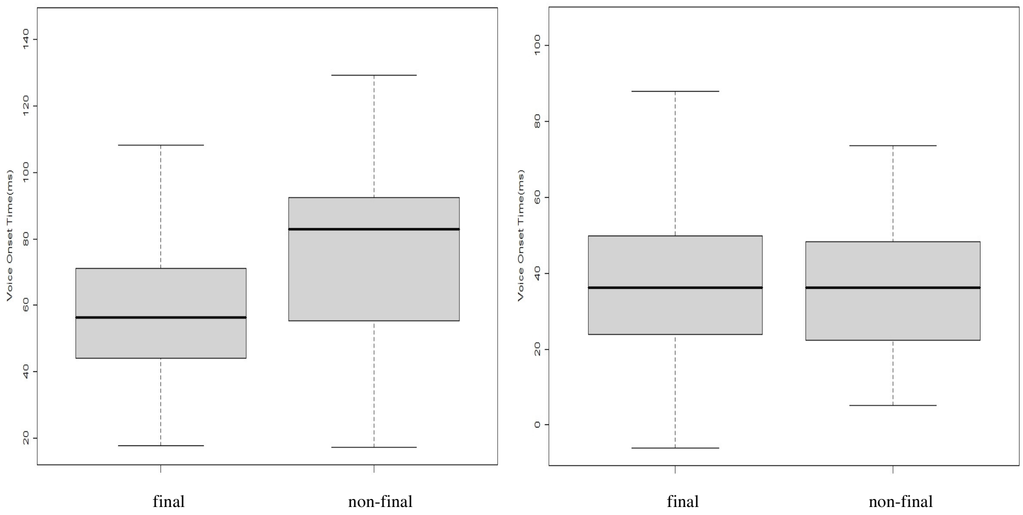

Figure 5. Average VOT of utterance-final and non-utterance-final tokens in F07 (left) and M08 (right)

	Previous R^2 (%)	New R^2 (%)
F07	11.8	19.11
M08	16.62	13.31

Table 5. Change in model performance after adding utterance position

3.6. Overall performance of the model

Speaker F07's final regression model has the following five variables: place of articulation, Buckeye word frequency, category of the previous phone, average speed of the three-word chunk centered at the target word, and utterance position. On the other hand, speaker M08's model ends up with four variables, place of articulation, Buckeye word frequency, category of the following phone, and duration of the following phone. The performance of the models after adding each variable is summarized in Table 6. In the final stage, speaker F07's model is able to account for 19.11% of the overall variation in VOT while speaker M08's model is able to account for 16.62% of the variation. In F07's model, the biggest increase in model performance happens when speech rate and utterance position are added; in M08's model, it happens when place of articulation and speech rate are added.

	Terms added	R^2 (%)
F07	Place of articulation	0
	Buckeye word frequency	1.7
	Category of the previous phone	1.9
	Average speed of the local 3-word chunk	11.8
	Utterance position	19.11
M08	Place of articulation	9.2
	Buckeye word frequency	9.6
	Category of the following	10.08
	phoneDuration of the next phone	16.62

Table 6. Summary of model performance of speaker F07 and M08

4. Discussion

In this study, we use naturalistic data from two speakers to model the variation of VOT in word-initial voiceless stops. The factors that are considered include place of articulation, word frequency, phonetic context, speech rate and utterance position. Overall, the following trends are observed, in at least one speaker's model: (a) VOT increases as the place of articulation moves from the lips to the velum; (b) higher frequency words have shorter VOT than lower frequency words, though the effect is not very strong; (c) when preceded by a vowel, VOT is shorter, and when followed by a vowel, VOT is longer, though the effect is weak; (d) in faster speech, VOT is shorter; (e) utterance-final stops have longer VOT.

The most interesting finding in the current study is that place of articulation is only shown to affect VOT in one speaker's model, but not in the other, though it has been claimed as a important conditioning factors of VOT in the literature. However, this doesn't mean that our results contest the canonical view, instead, our results show that in spontaneous speech, the VOT distinction among three stop categories can be overshadowed by other factors at play. In other words, for some speakers, the VOT distinction among [p], [t], and [k] is not strong enough to always be maintained. Our study also shows that word frequency, though widely reported as an important factor in pronunciation variation, doesn't have strong influence on VOT. In both speakers' model, there is a weak effect of word frequency that predicts shorter VOTs in higher-frequency words.

Speech rate is the only variable that has a strong effect on VOT in both speakers' models. Interestingly, the two speakers are shown to be sensitive to different speed measures. M08's data are best predicted by the most local speed measure, i.e. the duration of the following phone, while F07's data are best predicted by the medium local measure, i.e. the speed of the surrounding three-word chunk.

Even though we only examined two speakers' speech in the current study, they already show a wide range of individual differences. It is possible to attribute the differences to age and gender, since the older female speaker (F07) does have longer VOT than the young male speaker (M08), which is consistent with the results in the current literature. However, we think that a more important reason for the VOT difference is speech rate. As mentioned above, speaker F07 has the lowest average speech rate among all 19 speakers, and M08 has the highest, exceeding that of F07 by about 60%. In addition, our results also reveal individual differences in the variation pattern of VOT that can hardly be attributed to unknown differences in speech style. For one thing, speaker M08 shows a clear pattern of bilabial stops having the shortest VOT and velar stops having the longest one while in speaker F07's data, this pattern is not observed. It is not clear whether speaker F07 has no such distinction even in isolated word production, or the distinction is overshadowed by the various factors that are at play in spontaneous speech. In addition, utterance position is found to condition VOT values in F07, but not in M08, which indicates that the slower speaker slows down in utterance-final position while the fast speaker doesn't. This suggests that in addition to average speech rate, the variation in speech rate can also be an indicator of individual differences in speech style.

Altogether the model is only able to explain less than 20% of the variation in the data. One way to improve this model is to add more predictor variables. The literature on VOT and pronunciation variation has suggested a number of other factors that could potentially explain some of the remaining variation. These factors include contextual probability, prosody, the identity of the following vowel/consonant, neighborhood density and so on. (Note that disfluency is another factor that's not reported here. In fact, we did code cases for following disfluency, including (un)filled pauses and single word repetition, but the resulting division is very similar to that by utterance position.) It is possible that when these factors are considered, the performance of the model will be improved.

The other way to improve the performance is to use a different type of statistical model. As we know, linear regression models are limited to modeling linear relationship between the dependent variable and the independent variables. Therefore, it is inherently unable to model non-linear or non-homogeneous effects, which might exist in the VOT variation phenomenon. In addition, since all independent variables are internally coded as continuous variables, linear regression models also have difficulty in modeling the effect of categorical variables with more than two levels. Thus using a more general regression model might help explain more of the variation of VOT in spontaneous speech.

5. Concluding remarks

In this study, we present a first attempt to understand VOT variation in voiceless stops in spontaneous speech, and in particular, its relation with characteristics of the target word and the running context. Our results show that previously proposed factors, such as place of articulation and phonetic context, are still at play in spontaneous speech but the effect might be attenuated by the presence of other factors. We also show that lexical features, such as word class and word frequency, as well as contextual factors, such as speaking rate and utterance position, also have an effect on VOT, though the size of the effect is subject to individual differences. Finally, the overall low percentage of variation predicted by the linear regression model (despite the fact that we already excluded possibly non-homogeneous data) suggests that the actual variation pattern in spontaneous speech is highly complicated. In order to better model the variation phenomenon, more factors need to be considered and it might be necessary to use more complicated statistical tools.

References

Allen, J. Sean, Joanne L. Miller & David DeSteno. 2003. Individual talker differences in voice-onset-time. *Journal of Acoustic Society of America* 113(1), 544-552.

Baran, J.A., M.Z. Laufer & R. Daniloff. 1977. Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics* (5), 339-350.

Baum, Shari R. & Laurie Ryan. 1993. Rate of speech effects in aphasia: voice onset time. *Brain and Language* 44, 431-445.

Baayen, R.H., R. Piepenbrock & L. Gulikers. 1995. The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Bell, Alan, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory & Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113, 1001–1024.

Bell, Alan, Jason Brenier, Michelle Gregory, Cynthia Girand & Daniel Jurafsky. To appear. Predictability Effects on Durations of Content and Function Words in Conversational English. *Journal of Memory and Language*.

Bybee, Joan L. 2000. The phonology of the lexicon: evidence from lexical diffusion. In Barlow, M. & Kemmer, S. (Eds.), Usage-*based models of language*, 65-85. Stanford, CA: CSLI.

Byrd, Dani. 1993. 54,000 American stops. *UCLA Working Papers in Phonetics* 83, 97-116.

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

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

Fidelholz, James L. 1975. Word frequency and vowel reduction in English. In *Proceedings of CLS* 11, 200-213. Chicago:University of Chicago.

Forrest, Karen, Gary Weismer, & Greg S. Turner. 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.

Fosler-Lussier, Eric & Nelson Morgan. 2000. Effects of speaking rate and word frequency on conversational pronunciations. *Speech Communication* 29, 137-158.

Higgins, Maureen B., Ronald Netsell & Laura Schulte. 1998. Vowel related differences in laryngeal articulatory and phonatory function. *Journal of Speech, Language, and Hearing Research* 41, 712-724.

Hoit, Jeannette D., Nancy P. Solomon & Thomas J. Hixon. 1993. Effect of lung volume on voice onset time (VOT). *Journal of Speech, Language, and Hearing Research* 36, 516-520.

Koenig, Laura 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.

Kessinger, Rachel H. & Sheila E. Blumstein, S. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* 25(2), 143-168.

Kessinger, Rachel H. & Sheila E. Blumstein, S. 1998. Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics* 26, 117-128.

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

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

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

Neiman, Gary S., Richard J. Klich & Elain M. Shuey. 1983. Voice onset time in young and 70-year-old women. *Journal of Speech and Hearing Research* 26 (1), 118-23.

Petrosino, Linda, Roger D. Colcord, Karen B. Kurcz & Robert J. Yonker. 1993 Voice onset time of velar stop productions in aged speakers. *Perceptual and Motor Skills* 76 (1), 83-88.

Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume & Eric Fosler-Lussier. 2007. Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Pluymaekers, Mark, Miriam Ernestus & R. Harald Baayen. 2005. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118, 2561-2569.

Robb, Michael, Harvey Gilbert & Jay Lerman. 2005. Influence of gender and environmental setting on voice onset time. *Folia Phoniatrica et Logopaedica* 57(3), 125-133.

Ryalls, Jack, Marni Simon & Jerry Thomason. 2004. Voice Onset Time production in older Caucasian- and African-Americans. *Journal of Multilingual Communication*

Disorders 2 (1), 61 – 67.

Ryalls, Jack, Kristina Gustafson, & Celia Santini. 1999. Preliminary investigation of voice onset time production in persons with dysphagia. *Dysphagia* 14(3), 169-175.

Ryalls, John, Allison Zipprer, Penelope Baldauff. 1997. A preliminary investigation of the effects of gender and race on Voice Onset Time. *Journal of Speech and Hearing Research* 40(3), 642-645.

Schmidt, Anna M. & James E. Flege. 1996. Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. *Phonetica* 53(3), 162-179.

Sydal, Ann K. 1996. Acoustic variability in spontaneous conversational speech of American English talkers. Proceedings of *ICSLP '96*, 438-441.

Volaitis, Lydia E. & Joanne L. Miller. 1992. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America* 92(2), 723-735.

Whiteside, Sandra P., Luisa Henry & Rachel Dobbin. 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.

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

Whiteside, S. P. & Marshall, J. 2001. Developmental trends in voice onset time: some evidence for sex differences. *Phonetica* 58, 196-210.

Yao, Yao. 2008. An Exemplar-based Approach to Automatic Burst Detection in Spontaneous Speech. *Proceedings of CIL 18.*

Zue, Victor W. 1976. Acoustic characteristics of stop consonants: A controlled study. Sc. D. thesis. MIT, Cambridge, MA.

.