THE EFFECT OF SPEAKING RATE ON VOICE-ONSET-TIME IS TALKER-SPECIFIC

Rachel M. Theodore, Joanne L. Miller, and David DeSteno

Northeastern University
r.theodore@neu.edu

ABSTRACT

Talkers differ in phonetic properties of speech. One such property is voice-onset-time (VOT), an important marker of the voicing contrast in English stop consonants. Research has shown that VOT is affected by speaking rate: for any given talker, VOT increases as rate slows. The current work examines whether this contextual influence varies across talkers. Many tokens of /ti/ (Experiment 1) or /pi/ and /ki/ (Experiment 2) were elicited from talkers across a range of rates. VOT and syllable duration were measured for each token. The results showed that although VOT increased as rate slowed for all talkers, the extent of this increase varied significantly across talkers. For a given talker, however, the extent of the increase was stable across a change in place of articulation. These findings suggest that talker differences in phonetic properties of speech reflect talker-specific contextual influences.

Keywords: talker differences, speech production, VOT, speaking rate.

1. INTRODUCTION

Individual talkers differ in the detailed acoustic-phonetic information used to convey individual speech segments. These differences have been documented for both spectral [10,12] and temporal [2] properties of speech. Phonetic variation across talkers holds theoretical importance in terms of describing how listeners recover the segmental structure of language during comprehension. Early accounts of speech perception, which emphasized the abstract nature of linguistic representations [13], have recently been challenged by findings that indicate listeners retain in memory fine-grained information about how a talker implements specific speech sounds [5]. Other studies have demonstrated that familiarity with a particular talker's speech facilitates word recognition [11]. These findings support an alternative account of speech perception that views acoustic-phonetic variation across talkers as a useful source of information. Additional data on the ways in which talkers differ in speech production will facilitate the development of such an account.

Here we examine talker differences in speech production, focusing on VOT. VOT is a temporal property of stop consonants, measured acoustically as the latency between the release burst of the stop consonant and the onset of high-amplitude, periodic energy associated with the following vowel. In English, voiced stops are produced with short VOTs and voiceless stops are produced with longer VOTs [8]. Although this temporal contrast serves to distinguish voiced and voiceless stops relatively, the absolute VOT produced for a given stop consonant is robustly influenced by speaking rate: VOT increases as rate slows, especially for voiceless stops [6,14].

The current work examines whether the effect of speaking rate on VOT for a given voiceless stop is talker-specific (Experiment 1) and whether talker-specificity holds across place of articulation (Experiment 2).

2. EXPERIMENT 1

In Experiment 1, talkers were asked to produce the syllable /ti/ across a wide range of rates. In order to quantify the effect of rate on VOT, a linear function relating VOT to syllable duration was calculated for each talker. The slope of this function was used as an indicant of the effect of rate on VOT. Based on previous research, we expected all functions to have a positive slope [14]. The critical question was whether there is significant variability across talkers' slopes.

2.1. Subjects

Ten talkers (5 male, M1 – M5; 5 female, F1 – F5) were recruited. The talkers were native speakers of English between 18 and 31 years of age with no

known history of speech or hearing disorders. Talkers were either paid or received partial course credit for their participation.

2.2. Recordings

A magnitude-production procedure was used to elicit many repetitions of the syllable /ti/. In this procedure, talkers produce syllables at what they consider to be their normal rate and at rates relative to their normal rate. Each talker was recorded producing eight "runs" of syllables, with a run consisting of six repetitions of /ti/ at eight rates in the following order: normal, twice as fast as normal, four times as fast as normal, as fast as possible, normal, twice as slow as normal, four times as slow as normal, as slow as possible. All recordings took place in a sound-attenuated booth. Speech was recorded via microphone (AKG C460B) onto digital audiotape.

In total, 3840 syllables (8 runs X 8 rates X 6 repetitions X 10 talkers) were recorded. All recordings were digitized at a sampling rate of 20 kHz using the CSL system (KayPENTAX). Syllables produced at the normal rate at the beginning of each run were excluded from further analysis in order to help equate the number of tokens at each rate. In addition, the final repetition at each rate was excluded because this token may have been subject to a phrase-final lengthening effect [7]. Excluding these tokens left 2800 syllables (8 runs X 7 rates X 5 repetitions X 10 talkers) for acoustic analysis.

2.3. Acoustic Measurements

The Praat speech analysis software [3] was used to generate a waveform for each syllable. For each waveform, three durations were calculated: VOT, syllable duration, and vowel duration. VOT was calculated as the time between the onset of the release burst and voicing onset. Syllable duration, the primary metric of speaking rate, was calculated as the time between the onset of the release burst and voicing offset. Vowel duration, a secondary measure of speaking rate, was calculated as the time between voicing onset and voicing offset. All durations were calculated to the nearest millisecond (ms).

Of the 2800 syllables measured, 67 tokens (2.4%) were excluded from further analysis due to production anomalies or because a clear burst onset or voicing offset could not be determined. An additional 79 tokens (2.8%) whose syllable durations exceeded 799 ms were excluded because they were judged to be unnaturally long. As a result, measurements from 2654 tokens (94.8%) were used in subsequent statistical analyses. Both the number of tokens and the range of syllable durations were roughly comparable across talkers.

One trained experimenter conducted all acoustic measurements. To determine cross-experimenter reliability, a different trained experimenter measured one randomly selected run for each talker (approximately 13% of the tokens). Correlations (Pearson's r) between the two experimenters' measurements were 0.99 for both VOT and syllable duration. The mean absolute difference between the experimenters' measurements was 2 ms for VOT and 11 ms for syllable duration.

2.4. Results

For each of the 10 talkers, a linear function relating VOT to syllable duration was calculated using a least-squares prediction method. Table 1 shows the slopes of the individual talker functions, as well as the correlations between the functions and observed values as an index of goodness-of-fit. For ease of interpretation, slopes are given as change in VOT (ms) per 100 ms syllable duration.

Table 1: Slope [VOT (ms) / 100 ms syllable duration] and correlation (Pearson's r) for each talker's function.

Talker	Slope	r
M1	19	0.86
M2	21	0.88
M3	14	0.76
M4	8	0.74
M5	8	0.60
F1	16	0.78
F2	10	0.77
F3	21	0.82
F4	13	0.79
F5	11	0.80

In order to test the significance of the variability in talkers' slopes, a hierarchical linear modeling analysis (HLM, [4]) was applied to the data. For this analysis, the 2654 tokens were nested within each of the ten talkers as follows:

Level 1 model (group level):

$$\mathrm{VOT}_{ij} = \beta_{0j} + \beta_{1j}\,(\text{syllable duration}) + r_{ij}$$

Level 2 model (talker level):

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

With this structure, VOT is specified as a function of syllable duration, while incorporating the fact that sets of individual tokens are associated with specific talkers. Importantly, the Level-2 model allows the intercepts (β_{0j}) and slopes (β_{1j}) of the Level-1 model to vary across talkers. That is, the Level-2 model estimates the mean intercept (γ_{00}) and slope (γ_{10}) values across talkers while also testing if significant variability exists in these parameters (u_0 and u_1, respectively) as a function of stable talker differences. Results relevant to the current work showed that there was significant variability in the talkers' slopes [$\chi^2(9) = 489.42$, $p<.001$], indicating that the effect of syllable duration on VOT is not the same for all talkers[i].

3. EXPERIMENT 2

Results from Experiment 1 indicate that the effect of speaking rate on VOT is talker-specific. Experiment 2 extends this finding in two ways. First, we attempt to replicate this finding for the other two voiceless stop consonants in English, /p/ and /k/. Second, we examine whether the talker-specific rate effect is consistent across place of articulation. Talkers were asked to produce many repetitions of the syllables /pi/ and /ki/ across a range of rates using the procedure described for Experiment 1. For each talker, two functions were calculated relating VOT to syllable duration, one for /pi/ and one for /ki/. Based on the results from Experiment 1, we predicted that significant variability would be observed in the slopes of the talker functions within each place of articulation. The critical question was whether the slopes of the labial and velar functions would be the same for a given talker.

3.1. Subjects

Ten new talkers (5 male, M1 – M5; 5 female, F1 – F5) were recruited. The talkers were native speakers of English between 18 and 22 years of age with no known history of speech or hearing disorders. Talkers were either paid or received partial course credit for their participation.

3.2. Recordings

The magnitude-production procedure described for Experiment 1 was used to elicit repetitions of the syllables /pi/ and /ki/. As in Experiment 1, talkers produced eight runs of each syllable. The order of the labial and velar syllables was counter-balanced across talkers. All recordings followed the procedure outlined for Experiment 1.

In total, 7680 syllables (8 runs X 8 rates X 6 repetitions X 2 places of articulation X 10 talkers) were recorded. All recordings were digitized at a sampling rate of 20 kHZ using the CSL system. As in Experiment 1, syllables produced at the normal rate at the beginning of each run, as well as the final repetition at each rate, were excluded, leaving 5600 syllables (8 runs X 7 rates X 5 repetitions X 2 places of articulation X 10 talkers) for acoustic analysis.

3.3. Acoustic measurements

VOT, syllable duration, and vowel duration were measured for each syllable following the procedure outlined for Experiment 1. Of the 5600 syllables measured, 360 tokens (6.4%) were excluded from further analysis due to production anomalies or because a clear burst onset or voicing offset could not be determined. An additional 594 tokens (10.6%) whose syllable durations exceeded 799 ms were excluded because they were judged to be unnaturally long. As a result, measurements from 4646 tokens (83.0%) were used in subsequent statistical analyses.

Two trained experimenters conducted all measurements. To determine cross-experimenter reliability, a different trained experimenter measured one randomly determined run of /pi/ and /ki/ for each talker (approximately 13% of the tokens). Correlations (Pearson's r) between the two experimenters' measurements were 0.98 for VOT and 0.99 for syllable duration. The mean absolute difference between the experimenters' measurements was 4 ms for VOT and 29 ms for syllable duration.

3.4. Results

For each of the 10 talkers, two linear functions relating VOT to syllable duration were calculated, one for the labial syllables and one for the velar syllables. Table 2 shows the slopes of the individual labial and velar functions, correlations between the functions and observed values as an index of goodness-of-fit, and the difference between the labial and velar slopes.

To compare talkers' slopes within each place of articulation, two HLM analyses were performed following the model outlined for Experiment 1. Results showed significant variability in talkers' slopes for both the labial [$\chi^2(9) = 641.25$, $p<.001$]

and velar [$\chi^2(9) = 479.00$, $p<.001$] functions, replicating the findings from Experiment 1.

Table 2: Slope [VOT (ms) / 100 ms syllable duration] and correlation (Pearson's *r*) for the labial and velar functions for each talker, as well as the difference (Diff.) between the labial and velar slopes.

	Labial		Velar		
Talker	Slope	*r*	Slope	*r*	Diff.
M1	9	0.64	9	0.48	0
M2	5	0.55	10	0.65	-5
M3	21	0.88	18	0.87	3
M4	10	0.79	7	0.62	3
M5	4	0.48	5	0.55	-1
F1	12	0.64	17	0.64	-5
F2	10	0.81	9	0.76	1
F3	14	0.64	13	0.72	1
F4	4	0.39	6	0.64	-2
F5	21	0.65	24	0.77	-3

The main question was whether the slopes for /pi/ and /ki/ would be the same for a given talker. A paired-t test revealed that the mean difference between the labial and velar slopes (-0.80) was non-significant [$t(9) = -0.86$, $p=.41$]. However, this analysis is potentially misleading. The non-significant mean difference could indicate that for some talkers the labial slope was greater than the velar slope and for other talkers the labial slope was less than the velar slope, rather than indicating that the labial and velar slopes were the same for a given talker. In order to ensure that this was not the case, we conducted an additional HLM analysis nesting the talker-specific labial and velar slopes within talkers. No significant variability in the difference between the labial and velar slopes was found [$\chi^2(9) = 8.91$, $p>.50$]. These results indicate that the effect of syllable duration on VOT for a given talker is consistent across place of articulation.

4. CONCLUSIONS

The goal of the current work was to examine whether the effect of speaking rate on VOT varies across talkers. Results indicate that whereas VOT increases as rate slows for all talkers, the extent of this increase depends on who is speaking. Moreover, for a given talker, the extent of this increase is stable across place of articulation.

These findings have implications for how listeners accommodate talker differences in phonetic properties of speech. In terms of VOT, previous research indicates that listeners track VOT in a talker-dependent manner when rate is controlled [1]. However, talkers frequently alter their speaking rates [9]. The current results suggest that in order to accommodate fully for talker differences in VOT, listeners will have to take speaking rate into account.

5. REFERENCES

[1] Allen, J.S., Miller, J.L. 2004. Listener sensitivity to individual talker differences in voice-onset-time. *J. Acoust. Soc. Am.* 115, 3171-3183.

[2] Allen, J.S., Miller, J.L., DeSteno, D. 2003. Individual talker differences in voice-onset-time. *J. Acoust. Soc. Am.* 113, 544-552.

[3] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International*, 5, 341-345.

[4] Bryk, A.S., Raudenbush, S.W. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Newbury Park, CA: Sage.

[5] Goldinger, S.D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1166-1183.

[6] Kessinger, R.H., Blumstein, S.E. 1997. Effects of speaking rate on voice-onset-time in Thai, French, and English. *J. Phonetics*, 25, 143-168.

[7] Klatt, D.H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Am.* 59, 1208-1221.

[8] Lisker, L., Abramson, A.S. 1964. A cross-language study of voicing in initial stops: Acoustic measurements. *Word* 20, 384-422.

[9] Miller, J.L., Grosjean, F., Lomanto, C. 1984. Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica* 41, 215-225.

[10] Newman, R.S., Clouse, S.A., Burnham, J.L. 2001. The perceptual consequences of within-talker variability in fricative production. *J. Acoust. Soc. Am.* 109, 1181-1196.

[11] Nygaard, L.C., Pisoni, D.B. 1998. Talker-specific learning in speech perception. *Percept. Psychophys.* 60, 355-376.

[12] Peterson, G.E., Barney, H.L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175-184.

[13] Studdert-Kennedy, M. 1976. Speech perception. In: N.J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, 243-293.

[14] Volaitis, L.E., Miller, J.L. 1992. Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *J. Acoust. Soc. Am.* 92, 723-735.

[i] As a result of using syllable duration as the primary metric of speaking rate, VOT and speaking rate were not independent of one another. To ensure that our results did not rely on this coupling, all analyses reported in this paper were also conducted using vowel duration as the metric of speaking rate. The same pattern of results was found for both metrics.

[Supported by NIH DC 00130.]