Interactions between Speaking Rate and Temporal Implementation of Voicing Contrast: A Pilot Acoustic Study

著者(英)	Mayuki MATSUI
journal or publication title	NINJAL Research Papers
number	12
page range	47-62
year	2017-01
URL	http://doi.org/10.15084/00000853

国立国語研究所論集 (NINJAL Research Papers) 12: 47–62 (2017)
ISSN: 2186-134X print/2186-1358 online

Interactions between Speaking Rate and Temporal Implementation of Voicing Contrast: A Pilot Acoustic Study

MATSUI Mayuki

Postdoctoral Research Fellow, Theory & Typology Division, Research Department, NINJAL

Abstract

The current study examined how speaking rate affects phonetic implementation of phonological voicing contrast in word-initial stops, focusing on temporal properties such as voice-onset time (VOT), closure duration and vowel duration.

A pilot study of a Russian speaker showed that, the effect of speaking rate on VOT of voiced stops (negative VOT) is greater than that of voiceless stops (positive short-lag VOT), replicating previous studies. The results further revealed that (i) regardless of voicing specification, closure and vowel duration symmetrically decreases as speaking rate increases; (ii) the VOT/ vowel ratio and VOT/ closure ratio are roughly stable across speaking rates; (iii) rate-dependent change in negative VOT is predictable from the systematic change in closure duration.*

Key words: voicing contrast, stop, speaking rate, phonetics-phonology interface, Russian

1. Introduction

Phonology affects speech production to implement linguistic contrasts. At the same time, phonetic properties of speech sounds are highly variable as a function of other factors, such as speaking style (e.g., Smiljanić and Bradlow 2005, Maniwa et al. 2009, Igarashi et al. 2013), prosodic structure (e.g., Cho 2005, Cole et al. 2007, Mori et al. 2014, Erickson and Kawahara 2016), focus (e.g., De Jong and Zawaydeh 2002, De Jong 2004, Silbert and De Jong 2008), speaking rate (e.g., Summerfield 1981, Hirata 2004, Hirata and Whiton 2005) and so on. These studies have revealed that some phonetic properties are modified as a function of contextual factors,[1] while others remain stable. In other words, some properties show a context-dependent pattern, while others show a context-independent pattern. Context-dependent phonetic patterns can potentially be troublesome for language learners including children, since they must learn how to control articulators and how to perceive linguistic signals in a context-dependent manner.

Context-dependent phonetic patterns are troublesome not only for language learners but also for linguistic models of speech production and perception. In earlier models of phonetic implementation, it was hypothesized that detailed phonetic information, including context-dependent differences in the physical world, is abstracted away (Chomsky and Halle 1968, Pierrehumbert 1980, Keating 1990, see also Studdert-Kennedy 1976 for speech perception). In contrast, the

* I would like to thank the participant for her cooperation in the production experiment. I am grateful to the audience at NINJAL salon (June 2016) and the reviewer for feedback on this paper. I also wish to thank Sayaka Goto, Hyun Kyung Hwang, and Jaye Padgett for their insightful comments. Many thanks are also due to Donna Erickson for valuable input and proofreading. All remaining errors are my own.

[1] The term “context” here implies the context in a broader sense, including positional, segmental, pragmatic, prosodic and temporal contexts.

growing body of literature has shown that listeners retain context-dependent acoustic properties (e.g., Goldinger 1996, 1998), and that phonetic details play a role in speech processing (e.g., Hawkins and Nguyen 2003). How to integrate context-dependent patterns (particularly, "physical" timing patterns) into phonology is an ongoing theoretical debate (Port and Leary 2005: 931). Thus, interactions between contextual effects and phonetic implementation of phonological contrasts are still not well-known, and much remains to be learned. This paper addresses this general issue, with particular interest in the effects of speaking rate on voicing contrasts in stops. As will be discussed below, the effects of speaking rate interact with voicing contrasts in interesting ways.

Many languages, including Japanese, English, and Russian, have so-called voicing contrasts in phonology. Phonological contrast of voicing in stop consonants is typically described in terms of voice-onset time in many languages (Lisker and Abramson 1964). Voice-onset time (henceforth, VOT) refers to a time interval between two temporal events, that is, the release of oral closure (the reference point) and the onset of the vocal fold vibration. That is to say, a negative VOT implies that vocal fold vibration begins before the oral closure is released. Likewise, a positive VOT implies that vocal fold vibration begins after the oral closure is released. Also, the magnitude of VOT gives us insights about how much the glottal and oral gestures overlap or lag. According to Lisker and Abramson (1964), cross-linguistic variations of phonetic implementation of phonological voicing contrasts generally fall into three VOT categories: negative VOT, positive short-lag VOT, and positive long-lag VOT.

While VOT is known to be associated with phonological contrasts of voicing in stop consonants, it is also affected by various other factors, such as place of articulation (e.g., Cho and Ladefoged 1999), prosody (Cho 2005, Cole et al. 2007), speaking rate (e.g., Summerfield 1981, Miller et al. 1986, Kessinger and Blumstein 1997, Magloire and Green 1999), and so on. Since VOT is defined by glottal and oral temporal events, a fundamental question in phonetics is how this inter-gestural timing is affected by a systematic change in global temporal properties, namely, speaking rate. At the same time, an important question in phonology is how such contextual effects interact with phonological voicing contrasts. Thus, the purpose of the current study is to examine the effect of speaking rate on temporal implementation of voicing contrasts. The following section provides a background of the effect of speaking rate on voicing contrasts. Section 3 describes methodology. Section 4 presents the pilot acoustic study. Section 5 discusses the results of the pilot study, with some interim remarks.

2. Background

A number of studies have addressed the effect of speaking rate on VOT in a series of languages such as English (Summerfield 1981, Miller et al. 1986, Miller and Volaitis 1989, Kessinger and Blumstein 1997), Icelandic (Pind 1995), French (Kessinger and Blumstein 1997), Thai (Kessinger and Blumstein 1997), Spanish (Kessinger and Blumstein 1997, Magloire and Green 1999), Catalan (Solé and Estebas 2000), Swedish (Beckman et al. 2011), and Russian (Kulikov 2012, Matsui 2012). These studies demonstrated that not all VOTs are equally affected by speaking rate. For example, Summerfield (1981) observed that the VOTs of English voiceless stops (positive long-lag VOT) greatly change as a function of speaking rate, while those of voiced stops (positive short-lag VOT) were roughly stable across speaking rates.

Later cross-linguistic studies have revealed that the absolute value of VOT increases as the speaking rate slows, although such rate effects are robust for positive long-lag and negative VOTs while short-lag VOTs were minimally affected (e.g., Kessinger and Blumstein 1997 for English, Thai, and French; Pind 1995 for Icelandic; Magloire and Green 1999 for English and Spanish; and Kulikov 2012 and Matsui 2012 for Russian).

While rate effects on VOT have received considerable attention for decades, what is still not clear is the relationship between VOT and the other temporal patterns in stops. For example, closure duration of voiceless stops is usually longer than that of voiced stops (cf. Kingston and Diehl 1994), but is it symmetrically affected by speaking rate? Considering the fact that, by definition, the possible range of negative VOT is constrained by the closure duration of the stop as schematized in Figure 1, it may be the case that closure duration is affected by speaking rate, resulting in a change in negative VOT.

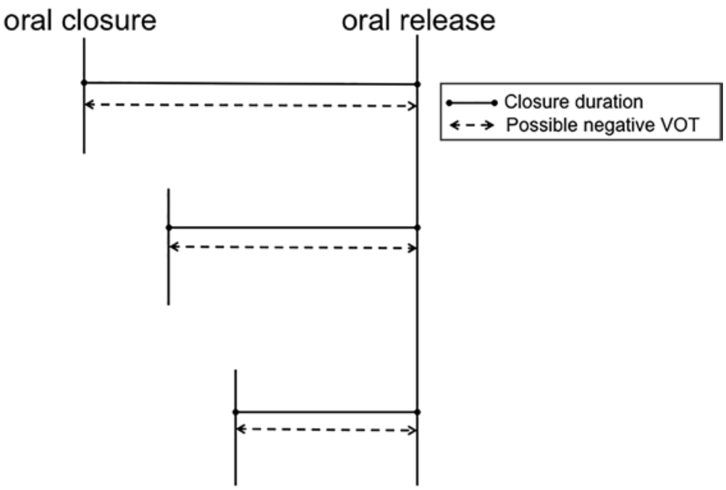

Figure 1. Schematic representation of the relationship between closure duration and negative VOT

Other than closure duration, Kessinger and Blumstein (1998) note a relationship between the vowel following word-initial voiceless stops (positive long-lag VOT) in English. They found that, as speaking rate increases, vowel duration also increases. They also revealed that the ratios between (positive long-lag) VOT and the following vowels are constant across speaking rates. However, Kessinger and Blumstein (1998) only tested the vowel following the voiceless stop having long-lag VOT. Therefore, it is not clear (i) whether the pattern is applicable to negative VOT and positive short-lag VOT, and (ii) whether voicing of the stop affects the following vowel duration. Thus, systematic analysis for both pairs of contrastive stops is needed. To summarize, while cumulative evidence suggests that short-lag VOT is cross-linguistically stable across speaking rates, the relation between VOT and other related acoustic properties at different speaking rates is understudied.

The current paper presents a pilot acoustic study of the effect of speaking rate on VOT, closure duration, and vowel duration in Russian. Russian is one of the languages that have a two-way voicing contrast. A previous acoustic study reports that over 97% of Russian so-called voiced stops have negative VOT, while so-called voiceless stops are realized with positive short-lag VOT (Ringen and Kulikov 2012). Considering previous studies of the rate effects on Russian VOT (Kulikov 2012, Matsui 2012) and cross-linguistic findings (e.g., Kessinger and Blumstein 1997), it would be predicted that in Russian the effects of speaking rate on VOT would be greater in voiced stops than in voiceless stops. In the current study, the results of the previous studies on VOT are replicated in a new dataset. Then, the effects of speaking rate on closure duration and vowel duration are examined. Finally, the relation between VOT and the other temporal properties are addressed. To the best of my knowledge, this is the first report on the effects of speaking rate on closure and vowel duration flanked by negative VOT or positive short-lag VOT. In the following sections, phonological transcriptions are given along with the latest IPA transcriptions of Russian (Yanushevskaya and Bunčić 2015).

3. Methods

3.1 Participant

The participant was a female native speaker of Russian (42 years old, from St. Petersburg) with no known history of speech and hearing disorders.

3.2 Speech materials

Speech materials consisted of three lexical (near-) minimal pairs contrasting voiced and voiceless bilabial non-palatalized stops (/p/, /b/) in word-initial position. The vowels following the initial target stops were either /a/, /o/, or /u/.

Table 1. Target words

Voiceless	Voiced
/par/ “steam”	/bar/ “bar”
/pot/ “swear”	/bot/ “boat”
/pusk/ “starting”	/buk/ “beech”

Each target word was embedded in a carrier sentence such as /skaʒi poʒalujsta, [target word], jeʃʲʃʲe raz/ (‘Please say [target word] once again’).

3.3 Procedures

The recording was conducted in a quiet room at the National Institute for Japanese Language and Linguistics. Speech materials were presented to the speaker in the Russian Cyrillic alphabet. Basic procedures followed *Magnitude production procedure* (Miller et al. 1986). The speaker read each sentence in four conditions: in the first condition, the speaker was asked to read the sentence at her comfortable, normal tempo; in the second condition, she was asked to read the same sentence twice as fast as her previous reading; in the third condition, she was again asked to read it at a normal tempo; finally, in the fourth condition, she was asked to read it at half the tempo of her previous reading. The second, the third, and the fourth conditions were defined as “*Fast*”, “*Middle*”, “*Slow*”, respectively. Each speech material was recorded five times in a random order.

3.4 Recording settings

Simultaneous audio and video inputs were recorded directly onto a Macintosh computer. Audio inputs were recorded via USB audio interface (Roland, UA-33) with a mono microphone (AKG, C520) at a 44.1-kHz sampling rate and 16-bit quantization level. Video inputs were recorded via an external webcam (Buffalo, BSWHD06M). These audio and video signals were stored together as a container format (.mov), from which an audio file (in a .wav format) was extracted for acoustic analyses.

3.5 Measurements

The utterances produced in *Fast*, *Middle*, and *Slow* conditions were annotated using the Praat speech analysis software (Boersma and Weenink 2010).[2] The focus of the current study is on (i) voice-onset time, (ii) closure duration, and (iii) vowel duration.

VOT was measured from the onset of the release burst (reflecting the onset of oral release) to the onset of quasi-periodic energy (reflecting the onset of vocal fold vibration). Closure duration was measured from the offset of the preceding phrase to the onset of the release burst. If a noticeable pause was inserted before the word-initial stop, such tokens were excluded from analysis. For vowel duration, the onset of the vowel was identified as the onset of the quasi-periodic wave occurring shortly after the stop burst. The offset of the vowel was identified based on an abrupt change in amplitude and the shape of the waveform. Sample waveforms and spectrograms are shown on Figures 2a and 2b.

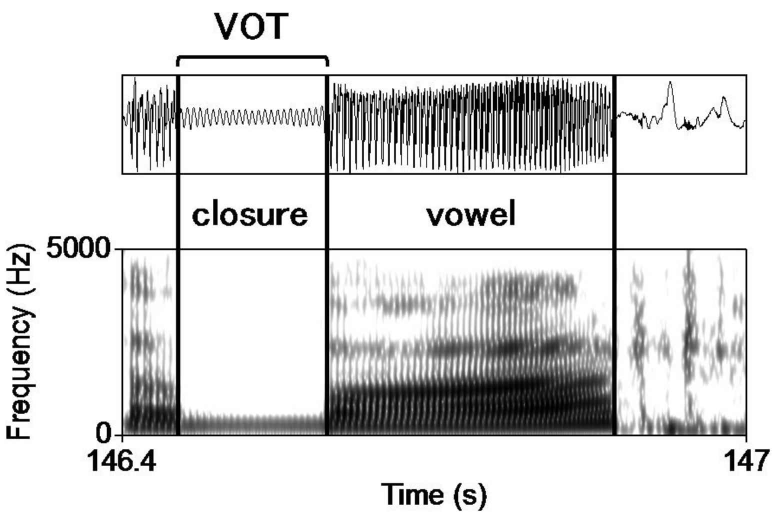

Figure 2a. Sample waveform and spectrogram of a voiced stop (/bar/, *Middle* condition)

[2] Video data were supplementarily used to double-check the closure timing of the lips.

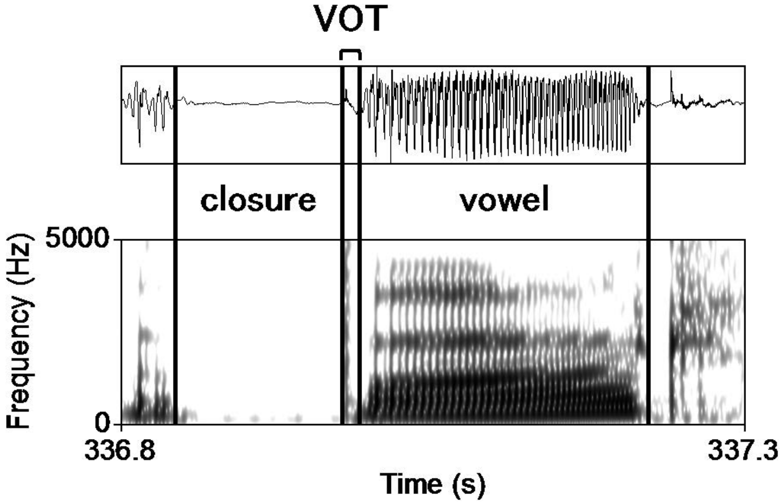

Figure 2b. Sample waveform and spectrogram of a voiceless stop (/par/, *Middle* condition)

In addition to these measurements, Consonant + Vowel (CV) duration of the target word (cf. Kessinger and Blumstein 1998) and the number of syllables per second were also considered as metrics of speaking rate. Following Kessinger and Blumstein (1998), CV duration was measured from oral release till the end of the following vowel.

An assumption here is that if our speaker manipulated her speaking rate successfully, CV duration should change as a function of speaking rate; that is, as speaking rate increases, CV duration should be shorter. More specifically, it is expected that if the rate manipulation was successful, CV durations at *Slow*, *Middle*, and *Fast* conditions would be: *Fast* < *Middle* < *Slow*.

As a reference, Figure 3 shows the mean CV duration in *Slow*, *Middle*, and *Fast* conditions. A two-way (rate condition × voicing specification of the stops) repeated measures Analysis of Variance (ANOVA) confirmed that there is a significant effect of speaking rate on CV duration [$F(2)= 41.702$, $p < 0.001$]. No interactions between speaking rate and voicing were significant [$F(2)= 0.006$, $p = 0.994$]. The three conditions were significantly different from each other [Tukey HSD for multiple comparison, *Slow* vs. *Middle*: $p < 0.001$; *Middle* vs. *Fast*: $p < 0.001$; *Fast* vs. *Slow*: $p < 0.001$]. In summary, as speaking rate increases, CV duration symmetrically decreases, regardless of voicing specification of word-initial stops. From these observations, it was confirmed that our speaker controlled the three speaking rates successfully.[3]

[3] This pattern was also confirmed with a traditional metric of speaking rate, such as number of syllables per second: as speaking rate increases, the number of syllables/sec increased. Since the patterns were consistent, and because of space limitations, only CV duration was considered in this paper.

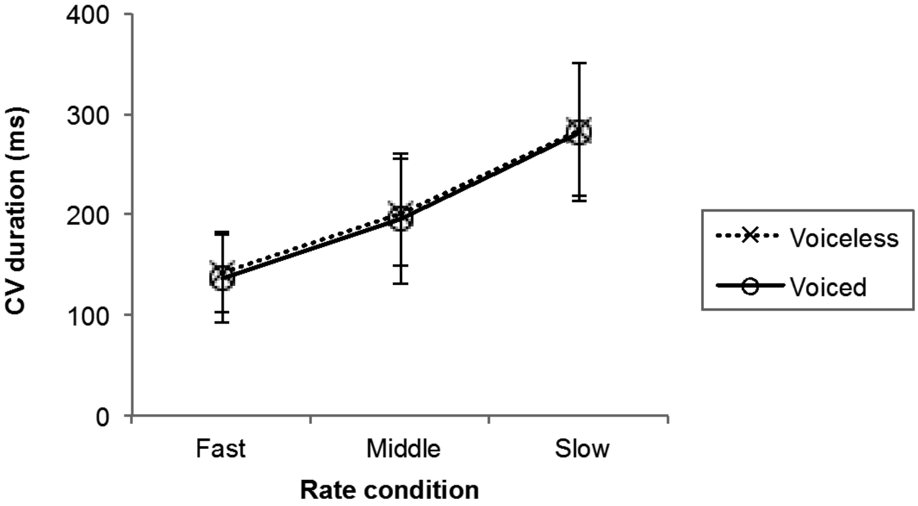

Figure 3. Mean CV duration broken down by rate condition. Error bars represent one standard deviation

3.6 Dataset

The total number of obtained tokens was 90 (6 sentences × 5 repetitions × 3 conditions). Out of 90, nine tokens were excluded due to undetectable release burst (N=1), presence of unnatural pause during an utterance (N=3), and discontinued voicing (N=5), where VOT could not be calculated since vocal fold vibrations would stop and start again during closure. The remaining 81 tokens were considered for further analyses.

4. Results

The results are divided into two parts. First, the effects of speaking rate on VOT, closure duration, and vowel duration are examined separately (section 4.1). Second, the relation between VOT and other measurements was investigated (section 4.2).

4.1 The effect of speaking rate

4.1.1 Voice-onset time

Since speaking rate was successfully manipulated (see Figure 3 above), a prediction is that the effect of speaking rate on VOT for voiced stops (negative VOT) will be greater than that for voiceless stops (positive short-lag VOT).

Figure 4 illustrates VOT as a function of CV duration. As is evident from Figure 4, the majority of voiced stops have negative VOT, whereas the majority of voiceless stops have positive short-lag VOT. This pattern is consistent with previous studies on VOT in Russian (Ringen and Kulikov 2012, Kulikov 2012). Moreover, importantly, even though the speaking rate was manipulated, there is almost no distributional overlap between voiced and voiceless stops along the continuum of VOT, which was also noted in the author's previous study with another dataset (Matsui 2012).

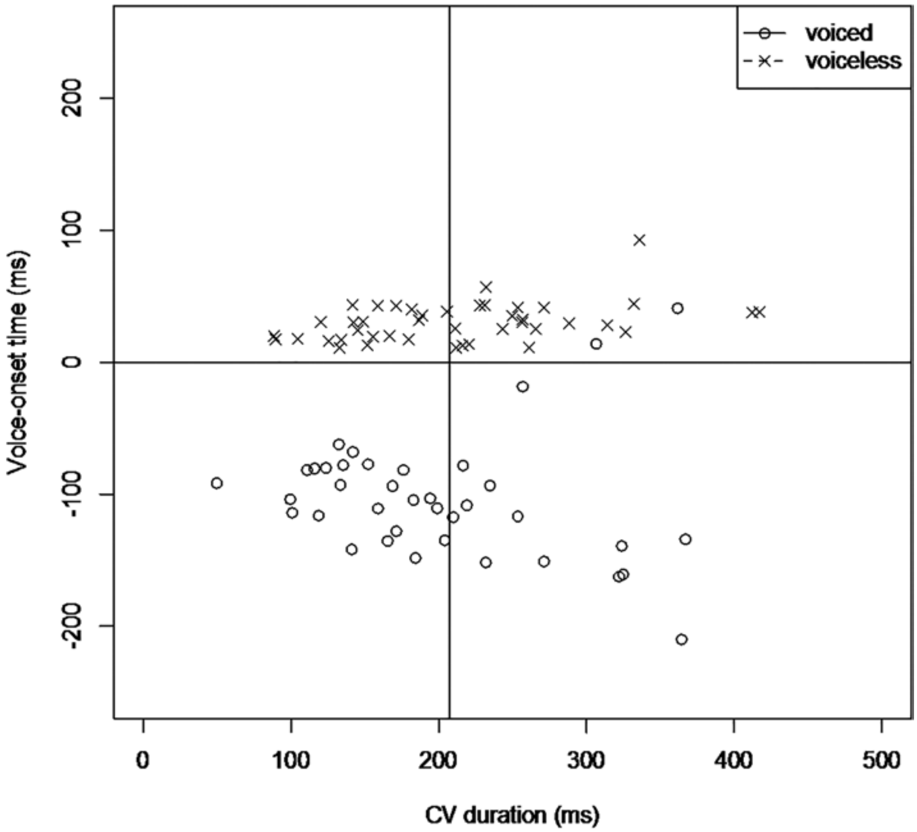

Figure 4. Voice-onset time as a function of CV duration (N=81). Vertical line represents the speaker's grand mean of CV duration (207 ms)

In order to further estimate the linear functions relating VOT with CV duration in both voiced and voiceless stops, a two-way Analysis of Covariance (ANCOVA) was run with VOT as a dependent variable. The predictors of VOT were CV duration (continuous variable) and voicing specification of stop (categorical variable). Before submitting the values to statistical analyses, outliers (N=3) were manually excluded. Moreover, in order to directly compare the difference in magnitude of the regression slopes, VOT values were converted into absolute values, as shown in Figure 5.

The results of the ANCOVA showed that there was a significant interaction between CV duration and voicing specification [$F(1) = 13.72, p < 0.001$]. Therefore, linear functions relating VOT with CV duration were estimated for voiced stops and voiceless stops separately. The ordinary least-square (OLS) regression analyses indicated that the regression slopes were significantly greater than zero for both voiced stops [$t(33) = 5.393, p < 0.001$; Adjusted $R^2 = 0.4524, p < 0.001$] and voiceless stops [$t(41) = 2.866, p < 0.001$; Adjusted $R^2 = 0.1466, p < 0.01$]. Importantly, estimated slope (β) for voiced stops ($\beta = 0.28526$) was approximately four times greater than that for voiceless stops ($\beta = 0.07747$).

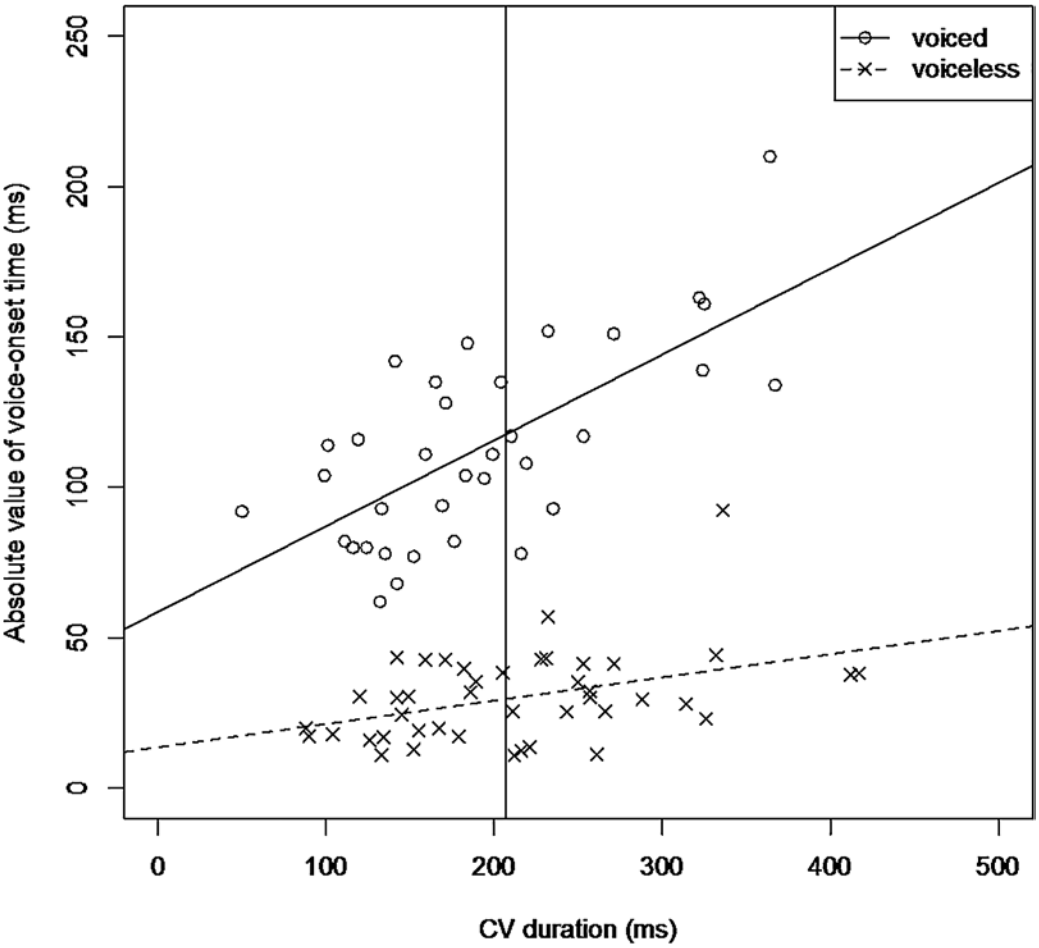

Figure 5. Voice-onset time (absolute value) as a function of CV duration (N=78). Intercepts are centered to the speaker's grand mean of CV duration (vertical line, 207 ms)

4.1.2 Closure duration

Figure 6 illustrates closure duration as a function of CV duration. In order to examine the rate-related change in closure duration, a two-way ANCOVA was again conducted with closure duration as a dependent variable. The predictors were CV duration and voicing specification of the stop.

The results showed that the effect of CV duration on closure duration was significant [$F(1)$ = 63.633, $p < 0.001$]. The effect of voicing specification was also significant [$F(1) = 7.078, p < 0.01$]. There was no significant interaction between the two predictors [$F(1) = 0.000, p = 0.99400$], suggesting that voiced and voiceless slopes may be similar to each other.

The OLS regression analyses indicated that, as can be seen in Figure 6, both slopes were significantly greater than zero for both voiced stops [$t(36) = 4.899, p < 0.001$; Adjusted $R^2 = 0.3833$, $p < 0.001$] and voiceless stops [$t(41) = 6.172, p < 0.001$; Adjusted $R^2 = 0.469, p < 0.001$]. The results further exhibited that in accordance with the results of the ANCOVA mentioned above, voiced and voiceless stops have highly similar slopes (β): β = 0.24579 for the voiced slope and β

= 0.2463 for the voiceless slope. It should be noted that the centered intercept of voiceless stops [130 ms, t (41) = 41.303, p < 0.001] was 14 ms greater than for voiced stops [116 ms, t (36) = 28.391, p < 0.001]. Taken together, this suggests that, while their distributions are different from each other, voiced stops and voiceless stops are symmetrically affected by speaking rate.

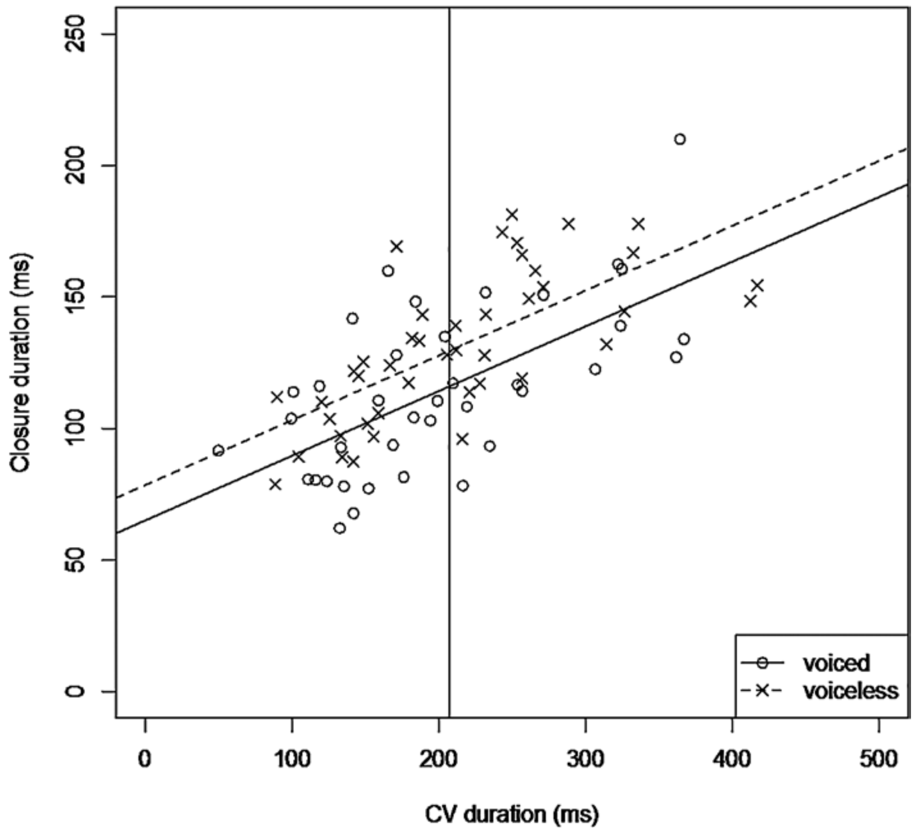

Figure 6. Closure duration as a function of CV duration (N=81). Intercepts are centered to the speaker's grand mean of CV duration (vertical line, 207 ms).

4.1.3 Vowel duration

Figure 7 illustrates vowel duration as a function of CV duration with regression lines. Again, a two-way ANCOVA was conducted with vowel duration as a dependent variable. The predictors of vowel duration were CV duration and voicing specification of the stop.

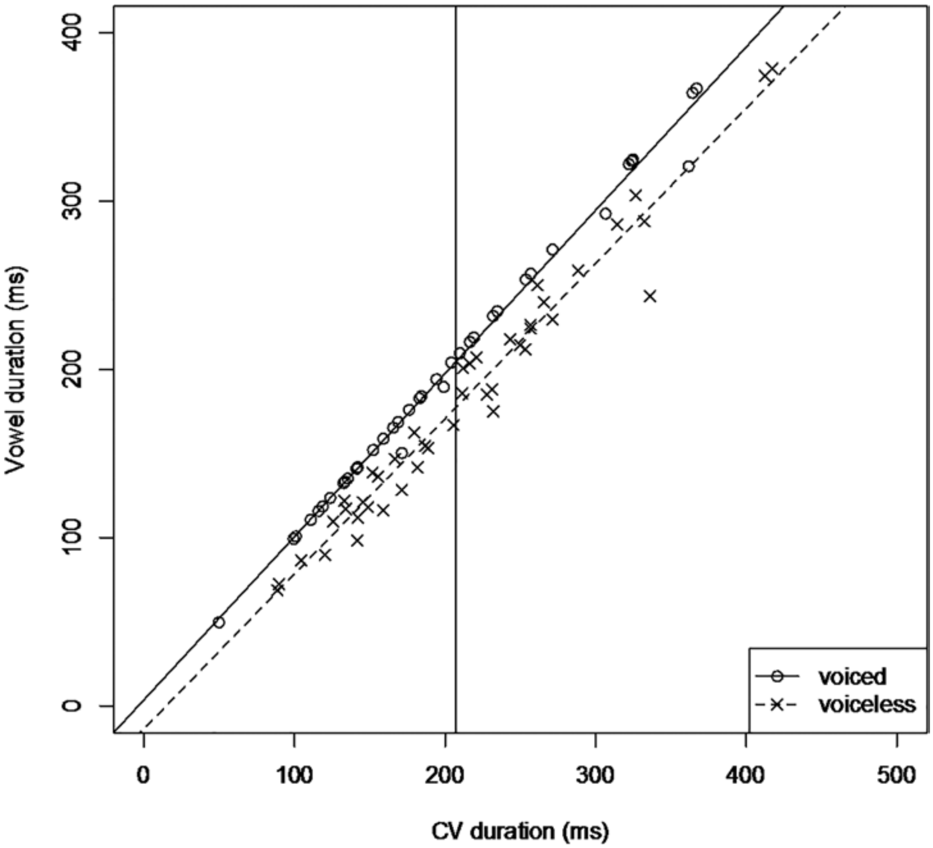

Figure 7. Vowel duration as a function of CV duration (N=81). Intercepts are centered to the speaker's grand mean of CV duration (vertical line, 207 ms)

The results showed that the effect of CV duration on vowel duration was significant [$F(1) = 3495.704$, $p < 0.001$]; as CV duration increases, vowel duration also increases. The effect of voicing specification was also significant [$F(1) = 115.792$, $p < 0.01$]; vowels following voiced stops are longer than those following voiceless stops. There was no significant interaction between the two predictors [$F(1) = 2.309$, $p = 0.133$], suggesting that the magnitude of the effects should be consistent regardless of voicing specification. The OLS regression analyses confirmed that the estimated voiced slopes [$\beta = 1.0$, $t(33) = 118.472$, $p < 0.001$; Adjusted $R^2 = 0.9976$, $p < 0.001$] are very similar to the voiceless slopes [$\beta = 0.9$, $t(41) = 33.989$, $p < 0.001$; Adjusted $R^2 = 0.9649$, $p < 0.05$]. It should be noted that the centered intercept of voiced stops [204ms, $t(38) = 169.77$, $p < 0.001$] was greater than that of voiceless stops [177ms, $t(41) = 83.27$, $p < 0.001$], suggesting that vowels are longer after voiced stops than after voiceless stops.

In summary, as CV duration increases, vowel duration increases for both voiced and voiceless stops. The vowels following voiced stops and voiceless stops are symmetrically affected by speaking rate.

4.1.4 Summary

The results of VOT, closure duration, and vowel duration demonstrated that all these measurements are affected by speaking rate: as speaking rate increases, VOT, closure duration, and vowel duration decrease. However, only for VOT, the magnitude of change in voiced stops (negative VOT) is significantly greater than that in voiceless stops (positive short-lag VOT), thus showing asymmetry.

4.2 The relationship between VOT and other temporal properties

In section 4.1, VOT, closure duration, and vowel duration were independently examined with CV duration as a mutual metric. This section takes a closer look at the relation between VOT and two other measurements.

Table 2 shows VOT/vowel ratios and Table 3 shows VOT/closure ratios at three rate conditions. As can be seen in these tables, these ratios are fairly stable across rate conditions. VOT/vowel ratio is 0.51–0.61 for voiced stops, and 0.15–0.20 for voiceless stops. VOT/closure ratio is roughly 1.0 for voiced stops and 0.22–0.24 for voiceless stops.

Table 2. VOT/vowel ratio (N=78)

Rate condition	Fast	Middle	Slow
Voiced			
Mean VOT (absolute value)	83	126	139
Mean vowel duration (ms)	136	195	271
Ratio	0.61	0.65	0.51
Voiceless			
Mean VOT (absolute value)	23	29	38
Mean vowel duration (ms)	119	173	247
Ratio	0.20	0.17	0.15

Table 3. VOT/closure ratio (N=78)

Rate condition	Fast	Middle	Slow
Voiced			
Mean VOT (absolute value)	83	126	139
Mean closure duration (ms)	83	128	139
Ratio	1.00	0.99	1.00
Voiceless			
Mean VOT (absolute value)	23	29	38
Mean closure duration (ms)	101	133	155
Ratio	0.23	0.22	0.24

5. Discussion

5.1 The effect of speaking rate on VOT, closure duration, and vowel duration

The results of VOT, closure duration, and vowel duration demonstrated that all these measurements are affected by speaking rate. For VOT, however, the rate effect on voiced stops (negative VOT) is roughly four times greater than on voiceless stops (positive short-lag VOT), replicating the findings of previous studies of Russian (Kulikov 2012, Matsui 2012) and with other languages (e.g., Kessinger and Blumstein 1997).

For closure duration, as speaking rate increases, closure duration decreases. As discussed in section 2, the possible range of negative VOT is constrained by the closure duration of the stop. The current study demonstrated that, regardless of the voicing specification of stops, closure duration is symmetrically affected by speaking rate, suggesting that rate-dependent change of negative VOT is predictable from the change in closure duration.

For vowel duration, Kessinger and Blumstain (1998) reported that in English vowel duration following word-initial /p/ (positive long-lag VOT) decreased as speaking rate increased. The current study extended this observation to negative and positive short-lag VOTs, at least for one speaker.

Taken together, the results revealed that the rate effect is asymmetric only for VOT in that it interacts with the voicing specification of stops. The magnitude of change in voiced stops (negative VOT) is significantly greater than that in voiceless stops (positive short-lag VOT). On the other hand, closure duration and vowel duration are symmetric, not interacting with voicing specification of stops.

5.2 Constant relationship between VOT and other measurements

The results showed that both VOT/vowel ratios and VOT/closure ratios are fairly stable across rate conditions. Of particular interest is the nearly perfect one-to-one relationship between negative VOT and closure duration. The results showed that (i) the closure duration is affected by speaking rate and (ii) the (negative) VOT/closure ratio is always roughly 1.0 (Tables 2 and 3). This means that, for voiced stops, the speaker almost always starts voicing simultaneously with lip closure. Based on these observations, it is suggested that at a gestural level, the timing of glottal constriction for vocal fold vibration strictly aligns with the oral closure onset. For the short-lag stops, on the other hand, it may be the case that the timing of glottal constriction for vocal fold vibration aligns with the oral release onset (cf. Solé 2007: 310, see also Benguerel et al. 1978 for French positive short-lag VOT). Assessing these possibilities is a topic for future investigation.

As for the relationship between VOT and the duration of the following vowel, Kessinger and Blumstein (1998) noted that the ratios between (positive long-lag) VOT and the following vowel are constant across speaking rates in English. This pilot study demonstrated that, at least for one speaker of Russian, (i) the pattern is applicable to negative and positive short-lag VOTs as well, and (ii) interactions between voicing contrasts of stops and speaking rate do not affect vowel duration.

5.3 Future perspectives

As a pilot study, the current study provided data from a single speaker. In order to confirm whether the patterns obtained in this study can be generalized to a group of speakers, further follow-up studies are necessary.

One possible direction for future studies is to discuss planned (controlled) vs. biomechanical properties of speech production. In the traditional (but admittedly still standard) modular view of phonetic implementation, phonetics is viewed as a translation from phonological output, implying that after the phonological output is determined; the remaining part is in the realm of speech biomechanics. However, increasing evidence suggests that some detailed phonetic properties in the speech signal are controlled by the speaker of a given language, while others are not controlled; the latter properties are biomechanical and governed by universal phonetic constraints (Kingston and Diehl 1994, Solé 2007, for a review). In this branch of study, it is claimed that some detailed phonetic properties are specified as input to speech production at a certain stage *after* phonological output but *prior to* speech biomechanics. They are then adjusted as a function of contextual factors such as speaking rate in order to achieve the intended linguistic contrast; others, however, are not inputted to production but executed at biomechanical levels. This view assumes a certain level of "controlled" phonetics *after* phonological output but *prior to* speech biomechanics, which is contrary to the traditional view of phonetic implementation.

The asymmetry in Russian VOT can be interpreted as a consequence of controlled vs. biomechanical implementations of VOT. That is, the VOT of (Russian) voiced stops changes greatly as a function of speaking rate, since the manifestation of negative VOT for voiced stops is part of the input to production, in other words, contextually controlled by speakers (of Russian). On the other hand, VOT of (Russian) voiceless stops is only minimally affected, since the manifestation of positive short-lag VOT is biomechanical and not controlled by speakers.[4] In order to further discuss the issues raised here, future studies should examine the effects of other contextual factors on VOT, such as the effect of prosodic boundary strength: if we assume that the manifestation of VOT of voiced stops is controlled at a certain stage *after* phonological output but *prior to* speech biomechanics inputted to production, VOT should also change as a function of other contextual factors.

In addition, in order to further examine the interactions between contextual factors and phonetic implementation of phonological contrasts, it would be insightful to directly compare languages with voicing contrasts to those without any voicing contrasts. To the best of my knowledge, previous studies have only examined languages with at least two or more phonological voicing contrasts. Therefore, testing the pattern of a language with *no* voicing contrast may forward our understanding of asymmetric rate effects.

References

Beckman, Jill, Pétur Helgason, Bob McMurray and Catherine Ringen (2011) Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics* 39: 39–49.

Benguerel, Adre-Pierre, Hajime Hirose, Masayuki Sawashima and Tatsujiro Ushijima (1978) Laryngeal control in French stop production: A fiberscopic, acoustic, and electromyographic study. *Folia Phoniat-*

[4] The discussion here is partly similar to the discussion in Beckman et al. (2011). Beckman et al. (2011) claimed that the phonological (underlying) laryngeal feature of stop consonants is privative in many languages on the basis of acoustic data: one is *phonologically* specified as [voice] and the other is unspecified, since negative VOT is "enhanced" as speaking rate decreases while positive short-lag VOT is not. The difference between the current discussion and Beckman et al. (2011) is the treatment of multiple acoustic cues. I do not attempt to discuss phonological feature specifications of Russian solely on the basis of VOT. What is under discussion here is not the issue of phonological specification but rather the stage *after* phonological specification.

rica 10: 81–100.

Boersma, Paul and David Weenink (2010) Praat: Doing phonetics by computer (Version 5.1.31). Retrieved April 4, 2010, http://www.praat.org/

Cho, Taehong (2005) Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English. *Journal of the Acoustical Society of America* 117(6): 3867–3878.

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

Chomsky, Noam and Morris Halle (1968) *The sound pattern of English*. New York: Harper and Row.

Cole, Jennifer, Heejin Kim, Hansook Choi, and Mark Hasegawa-Johnson (2007) Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* 35(2): 180–209.

De Jong, Kenneth (2004) Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics* 32(4): 493–516.

De Jong, Kenneth and Bushra Zawaydeh (2002) Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics* 30(1): 53–75.

Erickson, Donna and Shigeto Kawahara (2016) Articulatory correlates of metrical structure: Studying jaw displacement patterns. *Linguistic Vanguard* 2. Online publication. DOI 10.1515/lingvan-2015-0025.

Goldinger, Stephan D. (1996) Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(5): 1166–1183.

Goldinger, Stephan D. (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105: 251–279.

Hawkins, Sarah and Noël Nguyen (2003) Effects on word recognition of syllable-onset cues to syllable-coda voicing. In: John K. Local, Richard A. Ogden and Rosalind A. Temple (eds.) *Phonetic interpretation; Papers in laboratory phonology* VI, 38–57. Cambridge: Cambridge University Press.

Hirata, Yukari (2004) Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32(4): 565–589.

Hirata, Yukari and Jacob Whiton (2005) Effects of speaking rate on the single/geminate stop distinction in Japanese. *Journal of the Acoustical Society of America* 118(3): 1647–1660.

Igarashi, Yosuke, Ken'ya Nishikawa, Kuniyoshi Tanaka and Reiko Mazuka (2013) Phonological theory informs the analysis of intonational exaggeration in Japanese infant-directed speech. *Journal of the Acoustical Society of America* 134: 1283–1294.

Keating, Patricia A. (1990) Phonetic representations in a generative grammar. *Journal of Phonetics* 18: 321–334.

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

Kessinger, Rachel H. and Sheila E. Blumstein (1998) Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics* 26(2): 117–128.

Kingston, John and Randy Diehl (1994) Phonetic knowledge. *Language* 70(3): 419–454.

Kulikov, Vladimir (2012) Voicing and voice assimilation in Russian stops. Unpublished Ph.D. dissertation, The University of Iowa.

Lisker, Leigh and Arthur S. Abramson (1964) Cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 384–422.

Magloire, Joel and Kerry P. Green (1999) A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica* 56: 158–185.

Maniwa, Kazumi, Allard Jongman and Travis Wade (2009) Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America* 125(6): 3962–3973.

Matsui, Mayuki (2012) Asymmetric effects of speaking rate on voice-onset time: The case of Russian. *Proceedings of the 5th ISEL Conference of Experimental Linguistics*, 85–88.

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

Miller, Joanne L. and Lydia E. Volaitis (1989) Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics* 46: 505–512.

Mori, Yoko, Tomoko Hori and Donna Erickson (2014) Acoustic correlates of English rhythmic patterns for American vs. Japanese speakers. *Phonetica* 71: 83–108.
Pierrehumbert, Janet B. (1980) The phonology and phonetics of English intonation. Unpublished Ph.D. Dissertation, MIT.
Pind, Jörgen (1995) Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics* 57(3): 291–304.
Port, Robert and Adam Leary (2005) Against formal phonology. *Language* 81: 927–964.
Ringen, Catherine and Vladimir Kulikov (2012) Voicing in Russian stops: Crosslinguistic implications. *Journal of Slavic Linguistics* 20(2): 269–286.
Silbert, Noah and Kenneth De Jong (2008) Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America* 123: 2769–2779.
Smiljanić, Rajka and Ann R. Bradlow (2005) Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America* 118: 1677–1688.
Solé, Maria-Josep (2007) Controlled and mechanical properties in speech: A review of the literature. In: Maria-Josep Solé, Patrice Speeter Beddor and Manjari Ohala (eds.) *Experimental approaches to phonology*, 302–321. Oxford: Oxford University Press.
Solé, Maria-Josep and Eva Estebas (2000) Phonetic and phonological phenomena: V.O.T. A cross-language comparison. *Proceedings of the XVII AEDEAN Conference.* Vigo, Spain. 437–444.
Studdert-Kennedy, Michael (1976) Speech perception. In: Norman J. Lass (ed.) *Contemporary issues in experimental phonetics*, 243–293. New York: Academic Press.
Summerfield, Quentin (1981) Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7: 1074–1095.
Yanushevskaya, Irina and Daniel Bunčić (2015) Illustrations of the IPA: Russian. *Journal of the International Phonetic Association* 45 (2): 221–228.

発話速度と有声性対立に関わる時間特徴の間の相互作用：音響的側面の予備調査

松井真雪
国立国語研究所 研究系 理論･対照研究領域 プロジェクト研究員

要旨

本論文では，語頭閉鎖音の有声性の対立が発話速度の影響をどのように受けるのかを検討した。特に，Voice-onset time（VOT），閉鎖の持続時間，後続母音の持続時間といった，時間的な特徴の変動に焦点を当てた分析をおこなった。

ロシア語母語話者を対象とした小規模な予備調査から，以下の結果が得られた。まず，ロシア語の有声閉鎖音の VOT（negative VOT）は，無声閉鎖音の VOT（positive short-lag VOT）よりも，発話速度の影響を大きく受けることを示した。この結果は，先行研究の追試という位置づけになる。次に，新しい研究知見として，（i）有声性の音韻指定に関わらず，閉鎖音と後続母音の持続時間は対称的に変動すること，（ii）VOT と母音時間の時間比率，VOT と閉鎖時間の時間比率は発話速度を変えたとしてもおおよそ一定であること，（iii）Negative VOT の可変性は，閉鎖の持続時間の変動から予測可能であること，以上の 3 点を指摘した。

キーワード：有声性の対立，閉鎖音，発話速度，音声学と音韻論の関係論，ロシア語