University of Amsterdam
Faculty of Humanities

Effects of speech rate and elicitation methods on Polish Voice Onset Time

BA thesis
Linguistics

Diego Robert Ostoja-Kowalski
Student number: 13347543

Supervised by
Dr. Marijn van 't Veer

Faculty of Humanities
University of Amsterdam
The Netherlands
June 26, 2023

Abstract

The differences in pronunciation caused by using different elicitation methods in phonetic research is currently poorly understood. This study aims to start filling this knowledge gap by investigating whether the Voice Onset Time of Polish stop consonants can differ between the tasks of picture naming and word reading. In order to better understand this effect, it simultaneously investigates a similar Voice Onset Time difference between slow and fast speech, as this effect has been established in other languages for prevoiced and aspirated stops. It has been explained using the theory of Laryngeal Realism, in which only these two categories of Voice Onset Time are underlyingly laryngeally specified. Data collected from 18 Polish speakers living in Poland and in the Netherlands shows longer Voice Onset Time for Polish prevoiced stops elicited using picture naming, but the exact mechanism behind this result remains unknown. The study also finds the expected lengthening of Voiced Onset Time in slower speech for voiced stops. Additionally, it uncovers a lengthening effect for the voiceless series, which is not supposed to happen based on Laryngeal Realism and which is sometimes present in previous data. Several solutions are proposed, in which the voiceless series would need to be laryngeally specified. Another explanation posits that the effect is not in fact of a phonological nature. Further research suggestions into the nature of the Voice Onset Time changes are presented. The study also puts forward a hypothesis for the origin of the elicitation method effect.

Acknowledgements

I would like to thank my supervisor, Dr. Marijn van 't Veer, for his continued support and guidance. I am especially thankful for the encouragement to be braver with my conclusions and to say what I actually think about my data. I am grateful that you were part of my phonetics and phonology adventure over the past three years, and I hope this will continue in the future. Also, this thesis would not look as cool as it does without your subtle pushing me towards LaTeX over the years.

I also want to thank my partner, Dana, for always being there, both during my studies and while I was writing this thesis. Thank you for supporting me both when things were going well and during times of procrastination and helplessness. I still cannot believe I managed to finish this work, and it would have been so much harder without you. Рахмет, Даночка.

Huge thanks need to go to all my participants, too. This thesis would be truly impossible to write without your help, and I was amazed at how many of you wanted to participate. I am especially grateful to those I met during this research, thank you for your willingness to help a stranger and contribute to science. I also want to thank everybody who assisted me in the recruitment process – realizing I had a full schedule of experimental sessions planned was one of the most uplifting moments of the past three months.

I also need to thank my flatmates and my Linguistics colleagues. Being able to talk about the struggle connected to the thesis writing process always helped to relax. It was also great to have reasons to go out and get out of the world of phonetics and phonology for at least a while. It was a pleasure to meet you all, and I hope we keep in touch.

I want to thank everybody who led me towards professional linguistics over the years, I would not be where I am without those summer camp activities and the Linguistics Olympiads. The curiosity about languages has always been there, but you helped it grow.

Finally, I want to thank my Mom for her continued emotional and practical support throughout my studies. I would not have managed to live in a foreign country without your guidance. I am also grateful for supporting me in my choice of studies, even as what I was doing started to become less and less comprehensible to you. You were a crucial part of how my interest in linguistics began, I hope you like the results.

Contents

1 Introduction

This thesis aims to explore differences in pronunciation between two paradigms of eliciting speech, picture naming and word reading. It also investigates a possible link with better understood effects of speech rate on pronunciation. The phenomenon used to search for this link is Voice Onset Time variation in Polish, a language that was hitherto underresearched in the subfield of speech rate influence on Voice Onset Time. The results are analyzed primarily in the framework of Laryngeal Realism in order to assess the validity of its claims.

1.1 Previous studies of elicitation methods

When taking phonetic measurements, phoneticians use various elicitation methods to collect specific responses. The methods exhibit wide variation, from the most common and simplest picture naming and word reading tasks to answering questions about stories (Pająk & Baković, 2010) or providing specific verb forms in a complex visual setup (Mitterer, 2018). There also exists a separate line of research into eliciting more naturalistic spontaneous speech (see Baker and Hazan (2011) for an overview). As such, elicitation methods arguably constitute the practical core of many phonetic experiments. Despite that, there has been very little research into these methods' various effects on the elicited speech.

The few studies on the topic that were possible to find focus mostly on the contrast between picture naming and word reading. Most of them show effects that are easily attributable to the presence of orthography in the reading tasks. For instance, Lanwermeyer et al. (2019) investigated the German orthographic sequence ⟨-ig⟩, which is variously pronounced as [ɪk] or [ɪç]. They found that Austrian German speakers produced more stop realizations during reading tasks, most likely due to ⟨g⟩ most typically representing stop consonants in Standard German. Kharlamov (2014) meanwhile investigated the pronunciation of Russian voiced obstruents, which undergo word-final devoicing, but are still written with graphemes associated with voiced obstruents. In this study, he measured the number of glottal pulses that were present during the articulation of these consonants after vowels (as the voicing present in the vowel usually lasts longer and bleeds into the following voiceless sound). While the measured obstruents were still mostly voiceless, he showed that orthographically voiced stops had more glottal pulses when elicited using text instead of pictures or spoken sentences, which he interpreted as partial reversal of the devoicing.

However, Zhang et al. (2019) present an effect that does not have such a straightforward orthographic explanation. In their experiment, English speakers produced word-initial voiceless stops /p t k/ with longer Voice Onset Time in the word reading task compared to picture naming. While they did not perform statistical analyses on this factor and only provided one figure illustrating the effect, their result still merits some attention. Firstly, it could provide us a window into the exact processes occurring during lexical access, if the relationship between Voice Onset Time and different access pathways is better understood. Secondly, it could be compared to much better understood effects of speech rate on the Voice Onset Time of stop consonants, explained in the following section.

1.2 VOT and speech rate

Voice Onset Time (VOT) is a measure used to capture laryngeal contrasts in consonants, particularly plosives. It is often not enough to characterize all the differences between series of consonants, but it has been successfully used for a wide selection of the world's languages (Abramson & Whalen, 2017). While the exact methodological definitions of VOT vary, it is generally defined as the time difference between the onset of glottal pulsation and the release of the stop (Abramson & Whalen, 2017). The three most commonly selected categories, following the classification described by Beckman et al. (2011), are short-lag, aspirated and prevoiced (see Figure 1). In short-lag stops, the onset of voicing is very close to the release of the consonant, yielding VOTs around 0 ms. Aspirated stops have a larger gap between the two, and thus their VOTs are above 40 ms. For prevoiced stops the onset of voicing precedes the release of closure, and thus they are characterized by negative VOTs below -40 ms. Spoken languages often contrast two or more series of stops distinguishable in large part by their VOT ranges. The English contrast, at least in word-initial positions, is usually that between short-lag (typically described as "voiced") and aspirated stops (typically described as "voiceless"). Other languages can have different contrasts: French distinguishes short-lag and prevoiced stops, Thai distinguishes all three categories (Kessinger & Blumstein, 1997), and Central Standard Swedish contrasts only prevoiced stops and aspirates (Beckman et al., 2011).

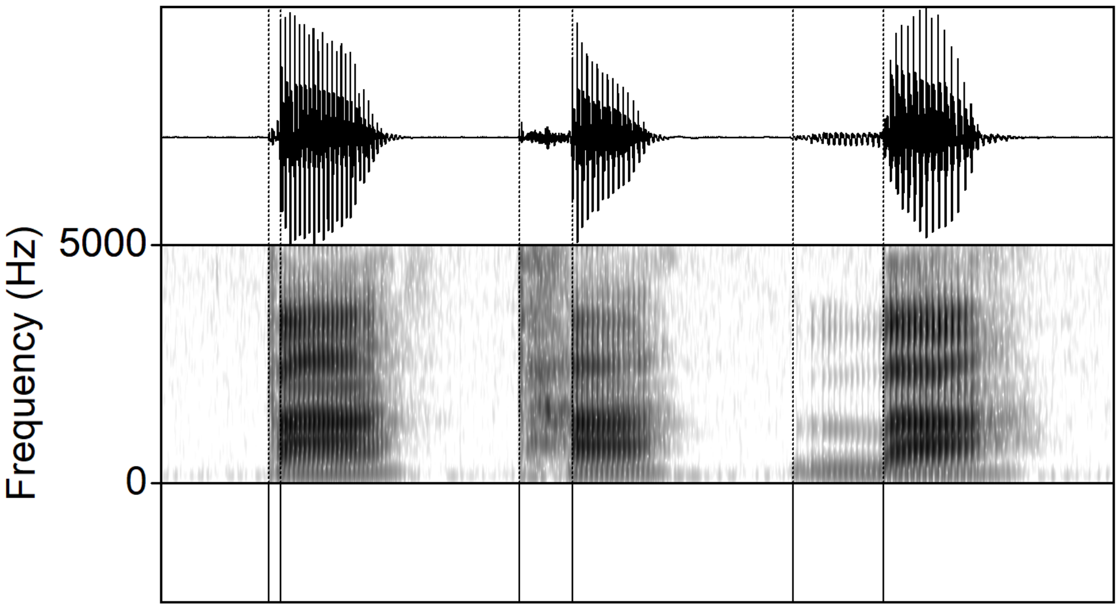


Figure 1: Examples of short-lag [pa], aspirated [pʰa] and prevoiced [ba], in this order, pronounced by the author. The vertical lines mark the release bursts and onsets of voicing. The VOT values of these tokens are 21 ms, 93 ms and -158 ms, respectively.

The most common findings are that slower speech rate is associated with more extreme VOTs for the prevoiced and aspirated categories (henceforth "longer VOTs"). In other words, the prevoicing and aspiration segments of stops become even longer, yielding VOT values further from 0. However, short-lag stops typically do not show significant changes in their VOT (see the introduction to Beckman et al., 2011; Lee and Berkson, 2019; Oh, 2009). As Zhang et al.'s (2019) findings are comparable with the speech rate effects, there is possibility that they have the same origin. It is also possible that their finding is simply the speech rate effect in disguise, as their study accounts for no measure of speech rate.

1.3 Laryngeal Realism and VOT

One explanation of the aforementioned VOT lengthening has been formulated in a framework called by some "Laryngeal Realism" (Honeybone, 2005). It is itself based on theories of phonology involving phonological features, such as the Distinctive Feature Theory (Hall, 2001). These theories propose that phonological segments of languages are bundles of features, and that each feature corresponds to a phonetic property of the segment. Laryngeal Realism has two further defining characteristics. Firstly, it rejects the use of binary features which would characterize e.g. the opposition /p/ : /b/ as [-voice] : [+voice]. Instead, it employs only privative features, using which /p/ : /b/ would be instead described as [∅] : [voice].

Importantly, it also rejects describing all two-way laryngeal contrasts with a single feature [voice], and proposes that some systems (e.g. English or German) should be instead characterized as based on other features. Firstly, their "voiceless" and "voiced" stops are often phonetically aspirated and short-lag, respectively, especially in word-initial positions. Additionally, Beckman et al. (2013) compare results on the voicing of Russian and German stop consonants in intervocalic positions (i.e. those most likely to show voicing) and found significant differences between the two languages. Russian voiced stops showed glottal pulsing for over 90% of their closure duration in 97% of all cases, while the corresponding figure for German "voiced" series is only 62.5%. Their conclusion is that Russian voicing is active and a realization of an underlying feature [voice], while the German voicing is passive and not reflective of any underlying feature.

Honeybone (2005) also discusses the so-called "Inner-German Consonant Weakening", a sound change which occurred in some varieties of German. It is a merger of the stops written as ⟨p t k⟩ into those usually represented using the letters ⟨b d g⟩. If it were interpreted as proper phonological voicing, this sound change would be typologically very unusual, since it would leave only the marked stop series intact. However, Laryngeal Realism provides the view that this is actually in line with the typological findings – as the German "voiced" stops are actually phonetically short-lag and are thus laryngeally unspecified, they are the unmarked series. Based on these and similar arguments, Laryngeal Realism proposes that the /p/ : /b/ contrasts of some languages (English, German) are underlyingly [spread glottis] : [∅] (Honeybone, 2005).

The explanation provided by Beckman et al. (2011) connects the lack of featural specification of short-lag stops to their unchanging VOT values. They posit that, when speaking more slowly, speakers will spend more time on realizing the phonetic cues of features, i.e. prevoicing for [voice] and aspiration for [spread glottis]. The observed lack of effect for short-lag stops then reflects the fact that they lack any laryngeal feature to be expressed. The authors note that this would make the Central Standard Swedish contrast overspecified as [spread glottis] : [voice], but they also conclude that this is a better explanation than other possible hypotheses. As an example, the lengthening of VOTs could be explained as "audience design" – as people speak slower, they have the time to make the consonants distinct enough so that there would be no confusion. However, Beckman et al. (2011) find that even in fast speech there was no VOT overlap between the two stop series, meaning there is no need to make them even more distinct in slow speech if the only reason being the VOT lengthening is intelligibility.

1.4 Polish laryngeal contrast

In light of this framework, Polish is an interesting language for further investigation. It possesses two series of obstruents distinguished by glottal activity, with a VOT contrast in stops (Jassem, 2003). As part of their study, Malisz and Żygis (2015) have measured the VOT values for both series and their data can be used to classify the Polish laryngeal distinction mostly as short-lag vs prevoiced (a minority of voiceless tokens falls into the aspirated range). Laryngeal Realism thus predicts only the voiced series to be phonologically specified, but other types of evidence in Polish have been interpreted to run counter to that.

For instance, both Schwartz et al.'s (2019) and Schwartz and Arndt's (2018) perception data show how lack of prevoicing is crucial to correct detection of Polish voiceless stops, while the voiced series can be easily identified when its VOT is manipulated. This leads Wojtkowiak and Schwartz (2022) to argue for the theory of Onset Prominence, in which all voiced stops, regardless of their phonetic realization, are laryngeally unspecified. The difference between e.g. German and French then does not come from their using different features, but it is hypothesized to stem from how exactly the voiceless stops are marked using the same feature [fortis][1].

Another theory based on evidence from Polish, often framed as an explicit response to Laryngeal Realism, is Cyran's (2014) Laryngeal Relativism. It also operates using privative features (its [H] and [L] features are largely equivalent to the previously discussed [spread glottis] and [voice]), with the difference that they do not have to always have the same phonetic realization between different languages. This is largely based on the phenomenon known as the "Cracow-Poznań sandhi voicing" (CPV) present in the dialects of Southern and Western Poland. In these dialects underlyingly voiceless word-final stops are often realized as the more marked voiced series when the next word begins with a sonorant, eg. *bra<u>t m</u>atki* 'mother's brother' [dm], *ja<u>k o</u>jciec* 'like father' [gɔ] (for more details see Strycharczuk (2012)). Cyran's (2014) explanation relies on the proposition that although in terms of VOT the dialects of Polish with and without VOT do not differ significantly, they are underlyingly different. He

[1] In the Onset Prominence theory phonemes are not simply bundles of features, they have internal structure and the internal placement of features can impact the phonetic realization.

posits that the CVP dialects /p/ : /b/ contrast is underlyingly [H] : [∅], identical to what we would find in e.g. English, and so the voicing is no longer marked – it is the default, unspecified state. The non-CVP dialects, meanwhile, would be specified as [∅] : [L], hence the "Relativism" in its name, there is no universal correspondence of underlying features and surface VOTs.

It is also worth examining the speech rate effects on Polish VOT, as they have not been studied in detail. The only data available comes from Malisz and Żygis (2015), who investigated the onsets of stressed syllables in four- and five-syllable long words, with word length serving as the measure of speech rate. Their results are largely inconclusive: in slower speech, the four-syllable words only showed longer voiceless VOT, and five-syllable words showed only lengthened voiced VOT. As such, more data is required to build a more detailed picture of the relationship between Polish VOT values and speech rate.

1.5 Current study

This study has two main aims. The first one is establishing whether there can be differences in the VOT of stop consonants elicited using different methodologies, which is primarily motivated by the scarcity of relevant data in the existing literature. The second one is confirming the predictions of Laryngeal Realism for the changes in Polish VOT depending on speech rate. There are two main reasons for investigating this effect. Firstly, there is more research into the phonological causes of this effect, and information about it could help understand the origins of the difference between elicitation methods. Secondly, some properties of Polish stop consonants have been used to refute the claims of Laryngeal Realism, but the speech rate effect has not yet been properly investigated in Polish. Such research could inform us better about the validity of that framework.

As such, the research questions are the following:

1. Are there differences in VOT between picture naming and word reading for Polish?
2. Which consonants (voiced or voiceless) have their VOT affected by speech rate in Polish and how?
3. Is there interaction between elicitation methods and speech rate for the VOT values in Polish?

Although the author of the current study does not fully subscribe to the theory of Laryngeal Realism, it is seemingly the only framework in which researchers have attempted to explain the VOT changes depending on speech rate. Due to that, the hypotheses will be formulated according to the theory, and the analysis will be primarily conducted to confirm its predictions. In light of Laryngeal Realism, speech rate is expected to affect only the voiced Polish stops, lengthening their VOTs in slow speech. The voiceless stops, as a consequence of being short-lag, should show no significant differences between the two speech rates. However, if the Polish laryngeal contrast is underlyingly different than what the theory predicts, it is likely that the voiceless stops will be affected. If the VOT difference between elicitation methods is not a by-product of the speech rate effect, it is expected to show up for at least one category of stops. Since Zhang et al.'s (2019) results show the effect for the aspirated stops in English, it is possible that elicitation methods only affect the stops with laryngeal specifications. As Laryngeals Realism expects only the voiced stops to have such specifications in the Polish language, they are expected to be more likely to exhibit this effect. Moreover, if it is caused by the same mechanism as the one behind speech rate effect, it is expected to show up for the same voicing category, possibly with some interaction between the two VOT effects.

There is also yet another aspect of this study that merits some explanation – the method for eliciting different speech rates. As explored by Lee and Berkson (2019), there exists a slight issue of the reliability of eliciting slow versus fast speech. Most studies to date have elicited faster speech by instructing participants to talk faster, but not so fast that the speech would become unintelligible. The authors of these studies note that this produces a wide range of variation in the quantities used to measure speech rate, e.g. word duration in Beckman et al. (2011) or syllable duration in Magloire and Green (1999). In order to avoid having to deal with this continuous variation, Lee and Berkson's (2019) study uses a metronome ticking at specific rates to help the participants control their speech rate. The current study also adopts this approach.

2 Materials and methods

The procedure described in this section, filed as FGW-885, was approved by the Ethics Committee of the Faculty of Humanities of the University of Amsterdam.

2.1 Participants

The participants were recruited both in Poland and in the Netherlands. The recruitment in Poland was conducted via the author's social circle in the city of Łódź. In total 13 participants were recruited from Łódź and the surrounding area. The recruitment in the Netherlands was conducted via Facebook, Instagram and Reddit posts, as well as flyers posted in Polish shops. This resulted in a total of eight participants being recruited from Amsterdam, The Hague and their surrounding areas.

Out of the total 21 participants, the data from 18 of them was analyzed. One participant found the experimental procedure too difficult and decided to stop the experiment. As a result, no data was collected from them. In the case of another participant, due to a mistake in the experimental procedure the recording equipment was not turned on during the second experimental stage. For the third discarded participant, their recording contained too much environmental noise, which prevented proper analysis of the recording. As a result, the analysis was performed on 11 participants living in Poland (one non-binary person, six women and four men) and seven participants living in the Netherlands (four women and three men). Their average age was 37.6 years (SD = 16.2 years), ranging from 20 to 79 years.

The only requirements for participation were being over the age of 16, having grown up using Polish at home and still using it in everyday life at least once per week. Before the experiment, the participants were shown the information brochure and the informed consent form and asked to confirm if they fit the selection criteria. Afterwards they were given a short questionnaire, where they self-reported their compatibility with the selection criteria. They were also asked about whether they were taught to play musical instruments – this piece of information was collected for a potential post-hoc analysis in case musical experience interacted with the use of a metronome in the study.

2.2 Stimuli

The study used two sets of stimuli. The first one was twelve bisyllabic Polish nouns, all of them describing easy to depict entities (e.g. *kościół* 'church', *banan* 'banana'). Each of them began with a different CV sequence, where C was one of the plosives /p t k b d g/ and V was one of the vowels /a ɔ/. The other four Polish vowel phonemes were not used for various reasons: there are too few bisyllabic nouns beginning with plosive + /u/ that are easy to depict, /ɨ/ does not occur after velars except in a handful of borrowings, and /ɛ i/ are rarely preceded by unpalatalized stops, mostly in borrowings. They were chosen to be bisyllabic as most Polish words have the stress on the second-to-last syllable, which provided a consistent prosodic context for all the stops – onsets of stressed syllables. In addition, twelve filler bisyllabic words beginning with other sounds were added. All of them exhibit the same stress pattern as the analyzed stimuli. Five additional words, fulfilling the same criteria as the filler words, were used for the practice stages (described in more detail in Section 2.3). The full set of stimuli, including fillers and practice items, can be found in Appendix A.

The second set of stimuli were line drawings corresponding to all the Polish words. Most drawings were selected from two standard sets of images by (Snodgrass & Vanderwart, 1980) and (Bates et al., 2003). However, an original drawing corresponding to the word *dama* 'queen (in card games)' had to be created. This was due to the fact that Polish happens to have very few bisyllabic words beginning in /da/, and none of them have any corresponding images in the two sets used for this study. Most of them were also deemed difficult to depict using a line drawing. As such, a decision was made to create a new line drawing for this word (as it is relatively easy to represent) and use it together with the other images. The two standard sets of images exhibit some significant internal stylistic variation, but the new drawing was deemed similar enough in style to most other pictures for the purposes of this research.

2.3 Equipment and procedure

The participants were randomly split into two groups. The difference between the groups was the set of stimuli used: one group was asked to read the word stimuli, the other one was being shown pictures and asked to name them.

For the convenience of participants, the experiments were conducted in their homes or other places of their choice, in quiet rooms without distractors. The recordings were made using a head-mounted unidirectional Samson QV microphone and a Marantz PMD620 MKII recorder, sampling at 44.1 kHz (16-bit). All the experiments were conducted fully using Polish by the author of this thesis. For each participant there were two experimental phases, each preceded by a practice stage. The microphone was recording only during the experimental phases.

During the whole experiment the participants were sat in front of the author's laptop. The stimuli were presented to them on the laptop's screen in a randomized order using PsychoPy (Peirce et al., 2019), with each stimulus appearing four times in each experimental phase. The stimuli were organized in four blocks of all 24 stimuli, with short breaks after each block. Most participants wore a single headphone earbud, over which they could hear a steady electronic metronome (also generated in PsychoPy). The other ear was left unobstructed so that they could hear themselves while speaking. One participant had to use their own headset due to health concerns. In their case the researcher made sure that the headset's noise cancelling function was turned off and that the participant could hear their own speech.

The experiment consisted of two experimental phases, and in each the metronome produced sounds[2] at a different rate, in order to elicit different speech rates. Every fourth beat sounded higher in order to signal when to start uttering the sentence. The metronome was set to 100 beats per minute in the slow condition and 150 beats per minute in the fast one. The first experimental phase always used the slower tempo.

The participants in the picture naming group were first familiarized with their set of stimuli on an additional brochure, in order to avoid eliciting alternative names as much as possible. Then, in both groups, the first practice stage began, where they were familiarized with the metronome and the production task. They were asked to embed the words represented by the stimuli in the carrier sentence _____ *jest tu na ekranie* ("_____ is here on the screen") and utter the sentences to the rhythm of the metronome in a trochaic pattern, with two syllables per beat (see Figure 2). They were also asked to take breaks and not say two sentences in a row. This was both to ensure that they do not run out of breath and that the exact beginning of prevoicing is possible to measure.

2.4 Acoustic analysis

The VOT lengths were later annotated manually in TextGrid files using Praat (Boersma & Weenink, 2023) and measured using a Praat script. Both the spectrogram and oscillogram were used to locate the points of release burst and the onset of voicing; the exact times were decided based on the oscillogram alone. The release burst was marked at the point where short aperiodic noise began, and the onset of voicing was located at the beginning of periodic glottal pulsing. The point of the release burst was automatically aligned to the zero of the waveform closest to the start of aperiodic noise (Figure 3A). The point of the onset of voicing was automatically aligned to the first positive-going zero during periodic glottal pulsing (Figure 3B).

In cases where the prevoicing was partial, i.e. there were breaks in glottal pulsing before the burst, the VOT was measured from the beginning of the first glottal pulsing phase,

[2]The sounds (short sine waves with frequencies of 900 and 1800 Hz) were extracted from the public domain recording *120BPMclicktrack.ogg*, created by the Wikimedia Commons user *Channel R*, available at https://commons.wikimedia.org/wiki/File:120BPMclicktrack.ogg [Retrieved April 23, 2023].

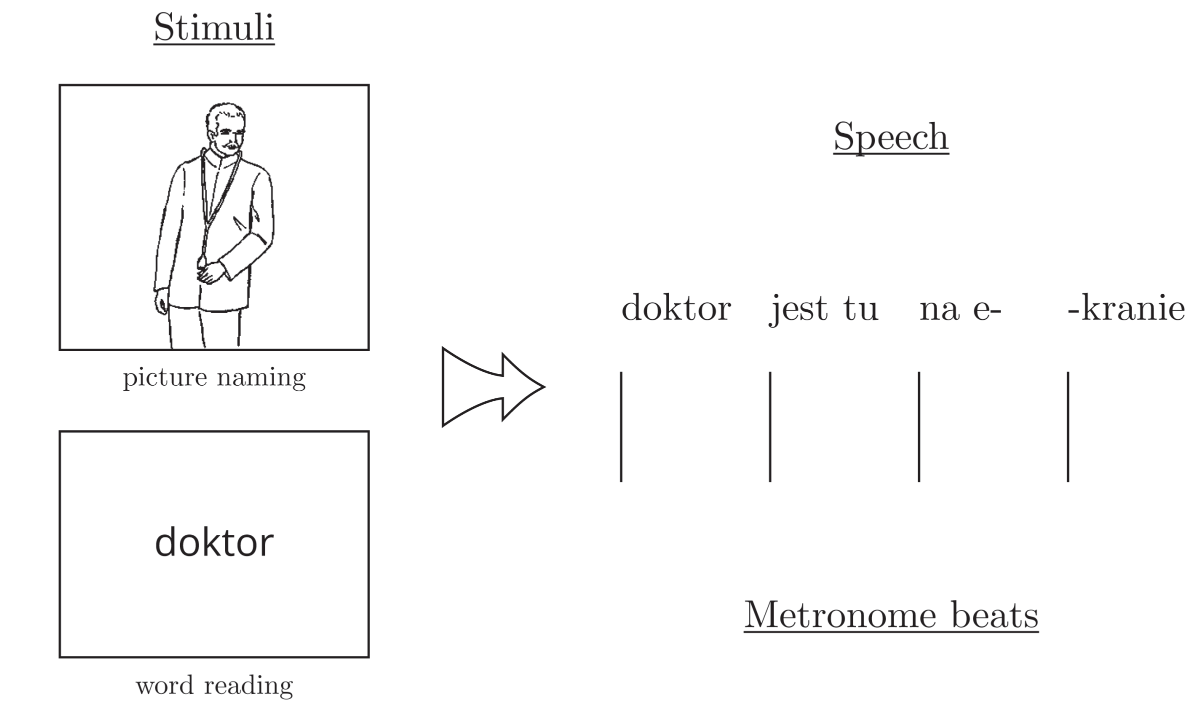


Figure 2: Schematic representation of the experiment. Left side – an exemplary pair of stimuli. Right side – intended utterance.

without accounting for the silent periods (see Figure 4). There was a total of 1728 elicited words, but some had to be discarded. One /b/ token was nasalized to [m], as such there was no observable release burst and the token was removed from the analysis. The same was done to one /k/ token which was followed by a devoiced vowel. As no glottal pulsing could be seen, the onset of voicing was impossible to mark. There were also eight instances, all elicited by the same speaker, in which the expected word *doktor* 'doctor' was replaced by its synonym *lekarz*. Another speaker replaced one token of the word *taśma* 'tape' by its synonym *kaseta*. All of these replacements were also removed from the analysis. As a result, a total of 1728 - (1 + 1 + 8 + 1) = 1717 different stop consonant tokens were annotated and measured for the analysis.

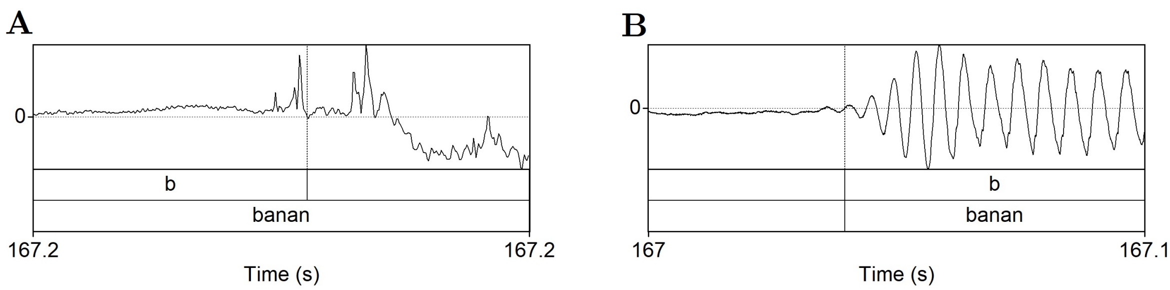


Figure 3: Examples of annotations of the release burst (A) and the onset of voicing (B).

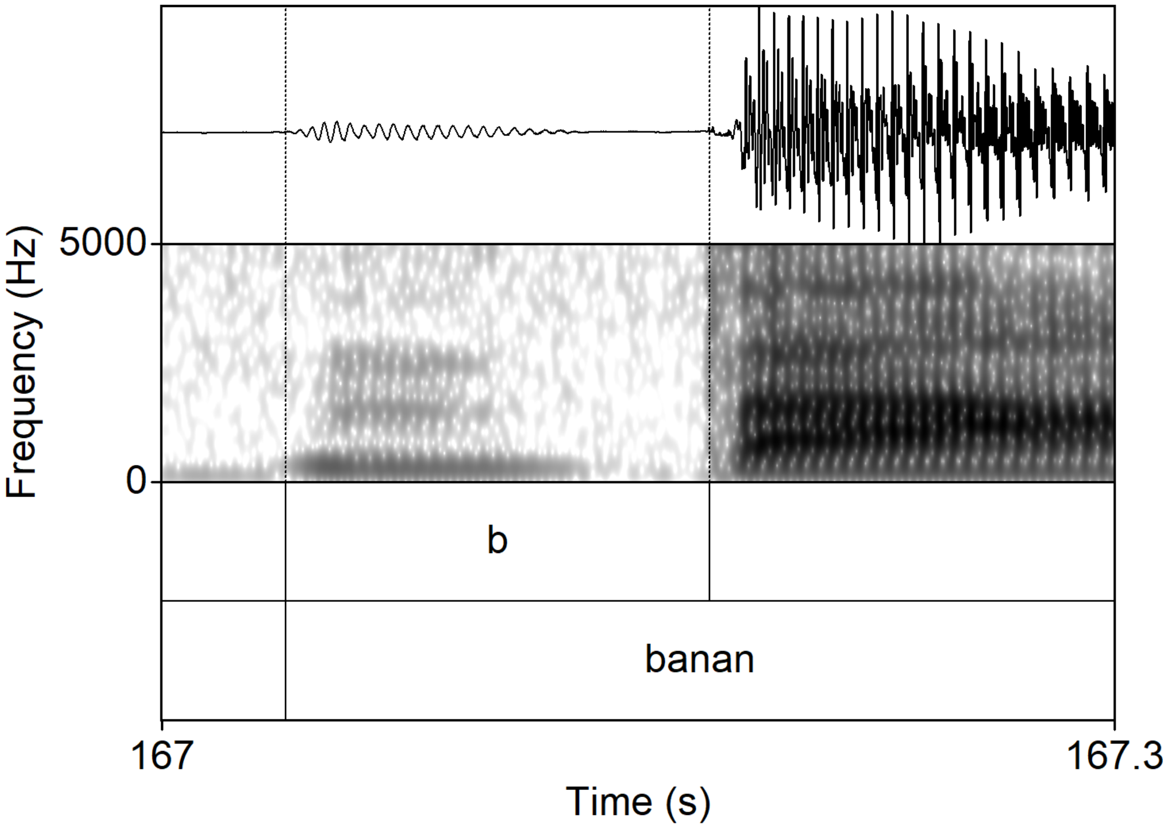


Figure 4: Example of a partially prevoiced /b/. Note the fading vibration in the oscillogram (top) and the disappearance of the bottom dark band in the spectrogram (middle).

3 Results

Figure 5 shows the entire distribution of VOT measurements for voiced and voiceless stops. The two series are largely separate, but a minority of voiced stops was pronounced with short-lag VOTs. This overlap is interesting, as previous Polish VOT data was reported without any overlap and with all voiced stops showing prevoicing (Malisz & Żygis, 2015). It is, however, similar to the overlap found by Magloire and Green (1999) for Spanish stops. Table 1 additionally displays the mean VOT for each stop consonant and each series. These data are consistent with the cross-linguistic patterns in VOT differences between places of articulation (Maddieson, 1997): going from labial to velar stops, average VOT values increase.

In order to answer the research questions, each series was separately analyzed in a four-way Speed (slow, fast) $\times$ Method (word reading, picture naming) $\times$ Place (labial, coronal, velar) $\times$ Vowel (/a/, /ɔ/) ANOVA. Section 3.1 presents the analysis of the voiced stops, while Section 3.2 contains the findings of the voiceless stop analysis.

3.1 Voiced stops

The first analysis was conducted on the subset containing voiced consonants (Figure 6). Firstly, in fast speech their mean VOT (M = -66.2 ms, SD = 41.0 ms) was shorter than

Table 1: Mean VOT values and standard deviations (in parentheses) in ms for all Polish stops. The final column is averaged over the entire laryngeal stop series.

	Labial	**Coronal**	**Velar**	**Total**
Voiceless	19.1 (10.2)	21.3 (8.9)	37.2 (11.2)	25.8 (13.0)
Voiced	-90.9 (47.5)	-76.5 (45.8)	-61.7 (50.3)	-76.4 (49.4)

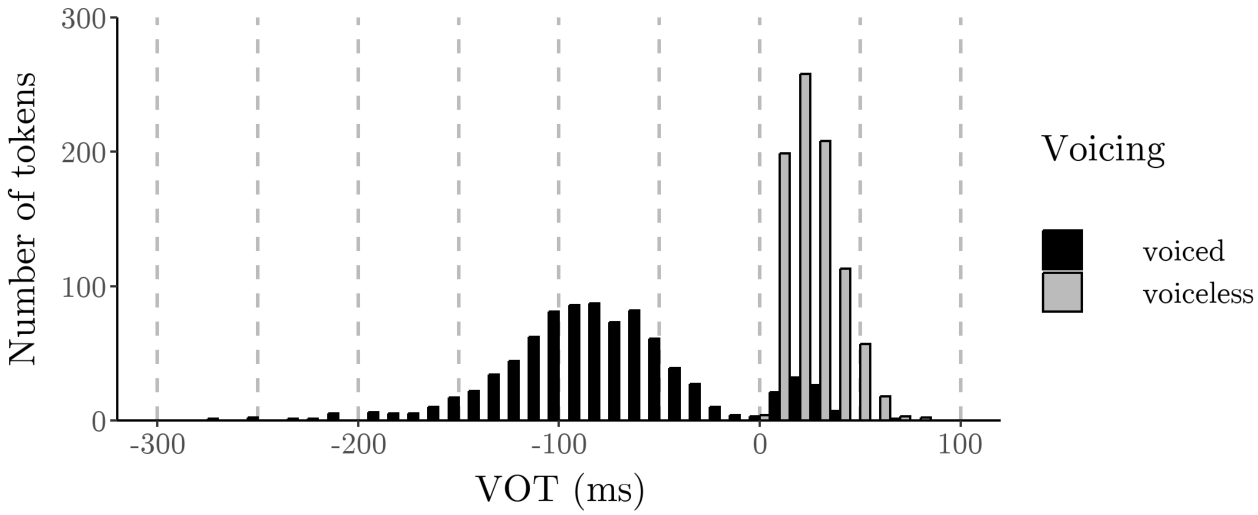


Figure 5: Overall distribution of measured Polish VOT values.
The bins are centered at 0, 10, 20 ms (etc.).

in slow speech (M = -86.5 ms, SD = 54.7 ms; $F(1, 831) = 41.40$, $p < 0.001$), in line with predictions of Laryngeal Realism. It can be also seen in the shift of the voiced series in Figure 8. There was no significant interaction of Speed with Method ($F < 1$), Place ($F(2, 831) = 1.92$, $p = 0.15$) or Vowel ($F < 1$). There were also no three-way or four-way interactions (all Fs < 1).

There was also a difference between the two elicitation methods, also showed in Figure 9. Voiced stops elicited using word reading had significantly shorter prevoicing (M = -66.0 ms, SD = 57.1 ms) than those elicited using picture naming (M = -84.8 ms, SD = 40.2 ms; $F(1, 831) = 35.55$, $p < 0.001$). This was an unexpected finding, given that Zhang et al. (2019) reported results suggesting that the text-based method would elicit longer VOTs. No interaction was found between Method and Place ($F(2, 831) = 1.35$, $p = 0.26$) or Vowel ($F < 1$), nor was there any significant interaction between all three of these variables ($F(2, 831) = 1.78$, $p = 0.17$).

In line with the contents of Table 1, place of articulation had a significant effect on VOT ($F(2, 831) = 28.93$, $p < 0.001$). Using unpaired t-tests, it was determined that the difference between /b/ and /d/ was significant ($T(564.93) = -3.69$, $p < 0.001$), as was the difference between /d/ and /g/ ($T(563.64) = -3.65$, $p < 0.001$). The analysis found no significant interaction between Place and Vowel ($F(2, 831) = 1.96$, $p = 0.14$). Finally, the VOT values did not differ significantly based on the vowel following voiced stops ($F < 1$).

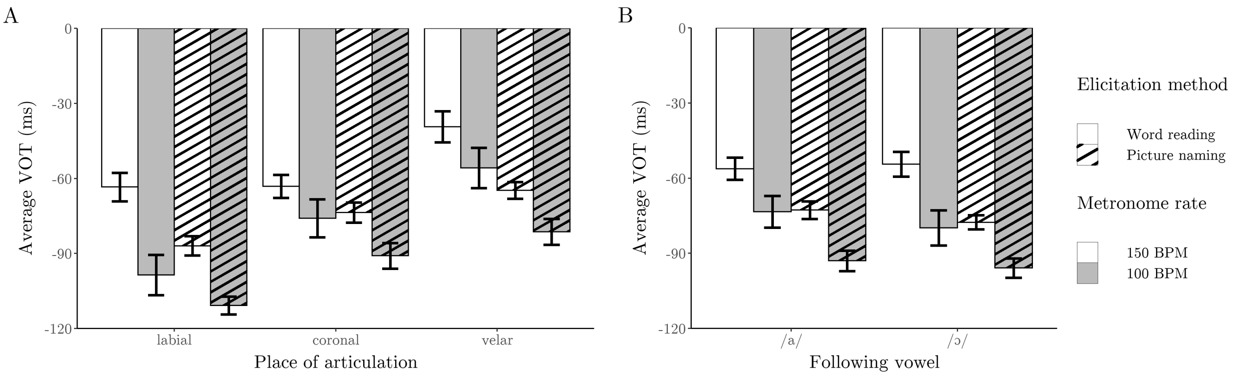


Figure 6: Effects of elicitation method and metronome rate on VOT for voiced stops by place of articulation (A) and by vowel (B). Error bars indicate standard errors.

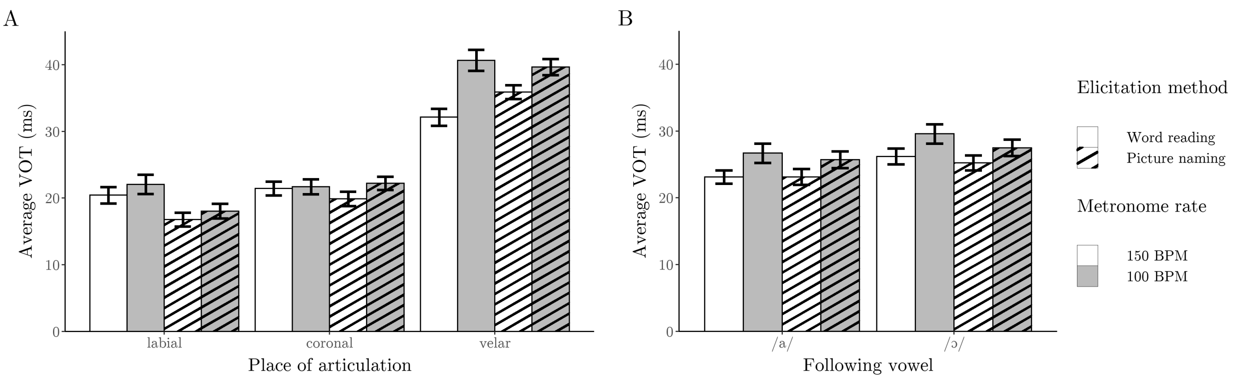


Figure 7: Effects of elicitation method and metronome rate on VOT for voiceless stops by place of articulation (A) and by vowel (B). Error bars indicate standard errors.

3.2 Voiceless stops

The second analysis concerned the voiceless stop tokens (see Figure 7). Similarly to voiced stops, their VOTs were shorter in the faster speech (M = 24.4 ms, SD = 11.9 ms) than when participants talked slower (M = 27.3 ms, SD = 13.9 ms; $F(1, 838) = 18.79$, $p < 0.001$). This finding stands in opposition to the behavior of short-lag stops described by Kessinger and Blumstein (1997) for English, French and Thai. It is, however, consistent with the description of English and Spanish short-lag stops in Magloire and Green (1999). The analysis did not find any significant interaction of Speed with Vowel ($F < 1$) or Method ($F < 1$). The Speed $\times$ Method $\times$ Place interaction was not significant ($F(2, 838) = 2.19$, $p = 0.11$), as were other three-way and four-way interactions (all Fs < 1). No reliable difference was detected in the VOT values between the two elicitation methods ($F(1, 838) = 2.19$, $p = 0.14$).

The three places of articulation differed significantly in terms of their average voiceless VOTs ($F(2, 838) = 283.29$, $p < 0.001$), as Table 1 suggested. Those effects were analyzed using t-tests analogously to the voiced stops, with differences found both between /p/ and /t/ ($T(559.79) = -2.76$, $p < 0.01$), as well as between /t/ and /k/ ($T(544.07) = -18.8$, $p < 0.001$). Place did not interact with Vowel ($F(2, 838) = 1.80$, $p = 0.17$) nor were there any three-way interactions, but interestingly it did interact both with Speed ($F(2, 838) = 4.84$, $p < 0.01$) and Method ($F(2, 838) = 5.01$, $p < 0.01$).

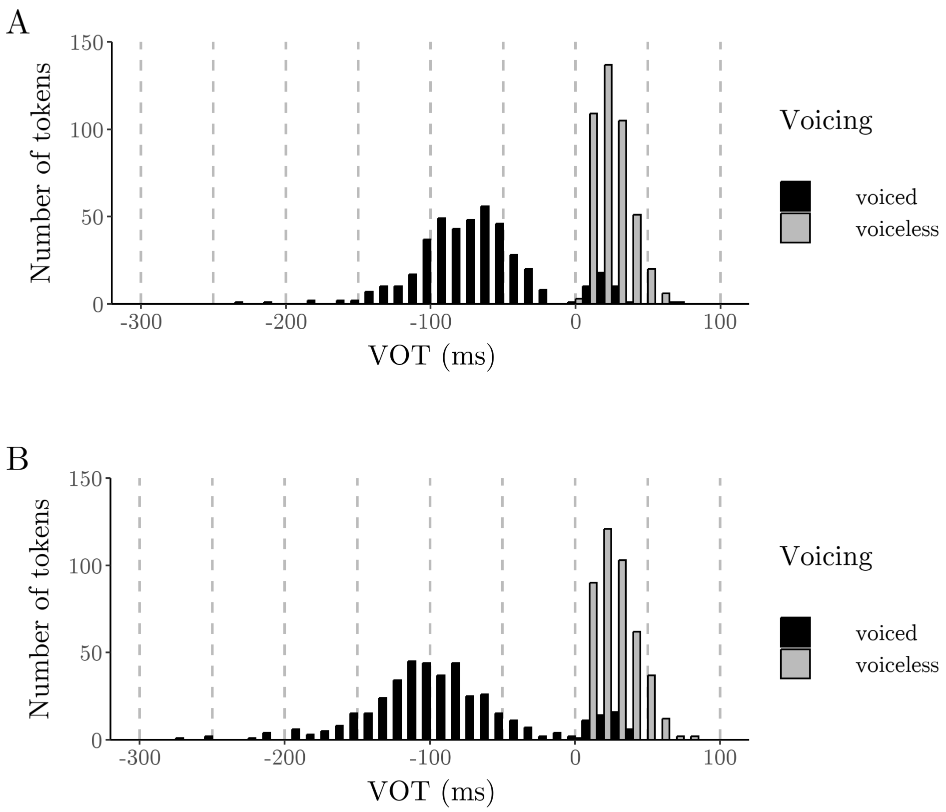


Figure 8: Distribution of VOTs in the fast (A) and slow (B) conditions. The bins are centered at 0, 10, 20 ms (etc.).

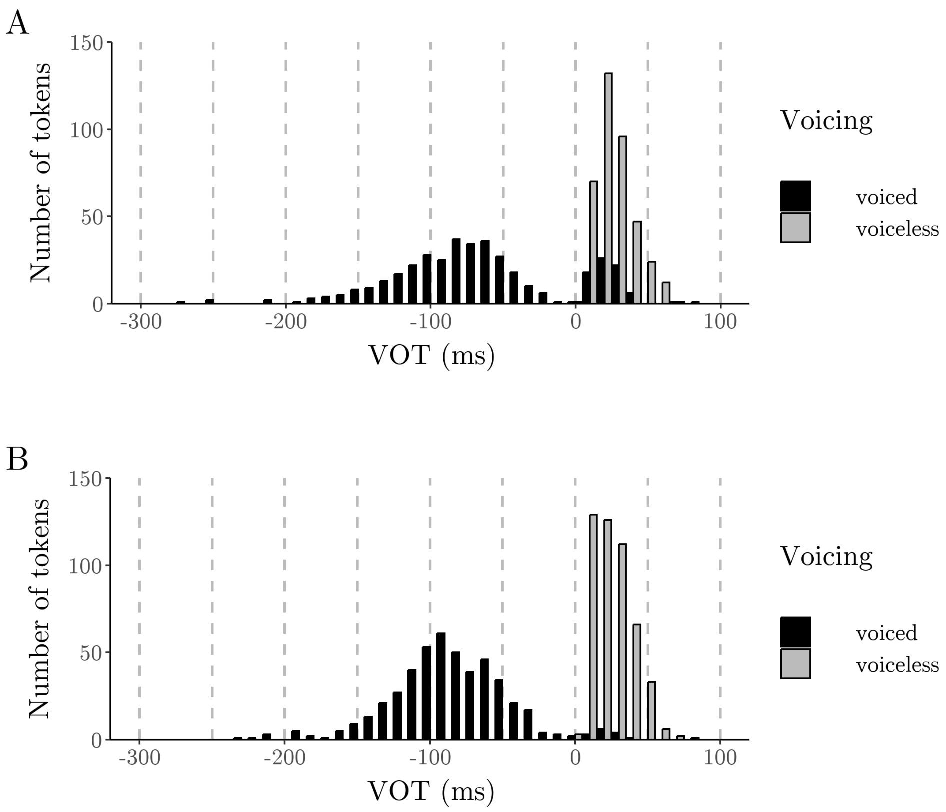


Figure 9: Distribution of VOTs during word reading (A) and picture naming (B). The bins are centered at 0, 10, 20 ms (etc.).

As illustrated by Table 2 and revealed by post-hoc t-tests, the difference between fast and slow speech was much greater for /k/ ($T(277.48) = -4.61$, $p < 0.001$) than for /p/ ($T(283.56) = -1.28$, $p = 0.20$) or for /t/ ($T(284.90) = -1.35$, $p = 0.18$). The interaction between Place and Method was meanwhile likely driven by the large differences exhibited by the /p/ phoneme ($T(261.65) = -3.22$, $p < 0.01$) compared to /t/ ($|T| < 1$) or /k/ ($T(248.15) = 1.01$, $p = 0.31$), as can be seen in Table 3.

Table 2: Mean VOT values and standard deviations (in parentheses) in ms for Polish voiceless stops, broken down by speech rate.

	/p/	**/t/**	**/k/**
Fast	18.3 (9.9)	20.6 (9.0)	34.2 (9.9)
Slow	19.8 (10.9)	22.0 (8.8)	40.1 (11.6)

Table 3: Mean VOT values and standard deviations (in parentheses) in ms for Polish voiceless stops, broken down by method of elicitation.

	/p/	**/t/**	**/k/**
Word reading	21.2 (10.7)	21.6 (8.5)	36.4 (12.2)
Picture naming	17.3 (9.9)	21.0 (9.2)	37.8 (10.3)

Unexpectedly there was a difference in VOT depending on what vowel followed the voiceless consonants. The vowel /a/ shows shorter VOT values for the preceding stops ($M = 24.6$ ms, SD = 13.0 ms) than /ɔ/ ($M = 27.0$ ms, SD = 12.9 ms; $F(1, 838) = 12.53$, $p < 0.001$). This result is difficult to interpret, but it is surprisingly similar to what was found by Beckman et al. (2011) in Central Standard Swedish.

4 Discussion

This section discusses the most important results presented in the previous section. It proposes several interpretations of the effect of speech rate on VOT of Polish stops, as well as stop consonants in general. Some of these interpretations are attempts at fitting the data to Beckman et al.'s (2011) argument in support of Laryngeal Realism, but others are rejections of this argument. The section also contains a shorter discussion of the difference in VOT induced by different elicitation methods, and the remaining problem of its connection to the speech rate effect.

4.1 Effects of speech rate

The first goal of this study was to check whether the VOT of Polish stops would change depending on the speech rate. The voiced stops behaved very similarly to what has been found for other languages possessing prevoiced stops. The slow metronome setting elicited

VOT values longer by 20.2 ms compared to faster speech, a result similar to the analogous figures for Central Standard Swedish (29.4 ms) and Spanish (23.4 ms for monolinguals, 28.8 ms for English-Spanish bilinguals) reported previously (Beckman et al., 2011; Magloire & Green, 1999). Based on previous arguments in favor of Laryngeal Realism, one could interpret this result as supporting the view that Polish voiced stops are phonologically specified for the feature [voice].

However, this research also unveiled a speech rate effect for voiceless stops, with a difference of 2.9 ms in VOT between fast and slow speech. Lack of VOT effects in short-lag stops is one of the core parts of Beckman et al.'s (2011) argument for Laryngeal Realism, and yet this data could be understood as evidence for short-lag stops being somehow phonologically specified. One possible solution to this would be to posit the existence of a VOT category existing between short-lag and aspirated stops. However, Magloire and Green (1999) results indicate the existence of a similar short-lag lengthening effect in slower speech not only in Spanish (13.5 ms for monolinguals, 2.1 ms for English-Spanish bilinguals), a language exhibiting a laryngeal contrast similar to the Polish one, but also in English (5.5 ms for monolinguals, 3.6 ms for English-Spanish bilinguals). Based on that data, they argue that the voiceless stops of languages similar to Spanish can be categorized as the same VOT class as short-lag stops of languages such as English. It is also interesting that no similar results were reported in Kessinger and Blumstein's (1997) study for Thai, French or English.

This issue of VOT lengthening effects in short-lag stops warrants further research. The first question would be how reliable these effects are, and whether they are replicable for any given language. The need for replicability comes from the inconsistent results present in English short-lag VOT data (the only language examined by more than one study in this subfield). This is problematic for anyone wanting to make definitive statements on whether short-lag stops change with speaking rate or not. The second question is whether the short-lag lengthening should be viewed as equally important as the corresponding effect in prevoiced and aspirated stops. If more data reveals it to be reliable, scholars need to decide whether this comparatively small VOT difference should be taken as evidence for short-lag stops being phonologically specified or not.

Another interpretation of the effect measured here is that Polish voiceless stops are indeed specified laryngeally as [spread glottis], and that the entire laryngeal system is overspecified. This would be similar to how Beckman et al interpret their evidence for Central Standard Swedish, but that explanation only introduces more questions. The primary issue stemming from it is to explain how this hypothetical Polish [spread glottis] feature can be suppressed so much that it is virtually indistiguishable from unspecified stops in other languages in terms of surface VOT values. While it does not seem likely, perhaps further research comparing multiple languages with different laryngeal contrasts could uncover some phonetic or articulatory features of e.g. Polish and Swedish voiceless stops that would reliably distinguish them from the voiceless stops in e.g. French. If so, we should also try to determine the mechanism which would be responsible for the suppression of aspiration in Polish voiceless stops (which would be expected under the assumption of [spread glottis] specification).

It might also be the case that there are two varieties of short-lag stops: some are underlyingly unspecified, while others are specified using some [fortis] feature distinct from both [voice] and [spread glottis]. Under that interpretation, Polish /p/ : /b/ would be specified as [fortis] : [voice]. The research recommendations would be the same as in the previous paragraph, with the addition that more languages need to be examined in order to determine whether such division of short-lag stops can be reliably made.

There also exists the possibility that the effects of speech rate on VOT do not in fact represent any underlying phonological facts, as they are more or less proportional to the sizes of the VOTs. It might be the case that the various phases of stop consonant production (or any speech sound in general) simply get contracted in a more or less uniform manner when we speak faster, and that the samples in previous studies were simply too small to reliably detect the small effect for short-lag stops. It would also mean that the VOT effects measured in this and previous studies should not be viewed as lengthening in slow speech, but as shortening in the faster setting.

One way to describe it more formally this would be to say that, after retrieving the relevant phonological specifications of speech sounds, the brain first creates an articulatory strategy expressible in some abstract units of time. This strategy would then get converted to a form that directly corresponds into the time intervals we observe both in the sound waveform and the in coordination of speech organs. Importantly, that conversion would take into account factors influencing how fast we need to speak in a given situation, and its output would depend on these factors. The VOT shortening effects would then be simply a result of that conversion, which would happen at a point where the phonological specification had already been processed and is not present in the input. It should thus be investigated how the durations of other phases of speech sound production vary with speech rate. If the time differences are proportional to the lengths of the corresponding intervals, that could constitute evidence for the non-phonological interpretation of VOT changes due to speech rate.

Unfortunately, it is difficult to interpret the data through any framework other than that of Laryngeal Realism. So far there have been no works at the level of Beckman et al. (2011) formulated in other theories which would address the speech rate effects on VOT. Even worse, existing rebuttals of Laryngeal Realism fail to engage properly with production data. Cyran (2014) treats "normal" voiced stops as equivalent in pronunciation to those resulting from CPV, even as Strycharczuk's (2012) data indicates that they differ significantly. For example, CPV-induced prevoiced stops (i.e. those resulting from being followed by a sonorant) had much shorter voicing durations than those created by voice assimilation to following voiced obstruents, suggesting they are not the same process. As his Laryngeal Relativism analysis relies on the two types of voiced stops being identical, this is a significant weakness of that theory.

Schwartz et al. (2019), meanwhile, point out what they see as a flaw of Laryngeal Realism in their production data: Polish voiced stops realized with only partial prevoicing have longer VOTs than fully prevoiced stops. They claim it is a contradiction, since longer VOTs would be stronger realizations of the feature [voice], but partial prevoicing would be its weaker realization, compared to continuous prevoicing. While there exist alternative interpretations[3],

[3]One such interpretation would be that partially prevoiced stops result when speakers aim for very long

consistent within Laryngeal Realism, it is also strange that the authors fail to engage with it using Onset Prominence. They simply claim that the contradiction can be avoided if assume there is no feature [voice] (as Onset Prominence proposes), and they proceed to ignore this effect, without providing any explanations for what its cause would be.

Given this context, we need a theory addressing both production and perception of VOT and other laryngeal phenomena, perhaps one with a more explicit model of how phonological representations correspond to how speech is produced and interpreted. Explicitness could especially be useful when trying to model whether the speech rate effect is an actual phonological phenomenon, or merely a by-product of late-stage articulatory planning.

4.2 Effects of elicitation methods

The second goal of this thesis was to determine whether using different elicitation methods would result in recording different VOT values. Based on Zhang et al.'s (2019) data, it was expected that voiced stops elicited using picture naming would have shorter VOTs than those produced during word reading. Contrary to that, this study found that the picture naming group produced voiced stops with VOTs longer by 18.8 ms compared to word reading. This result is difficult to interpret, but at least it shows that the choice of experimental methods can affect the phonetics of elicited speech. Further research should investigate possible causes of this effect, particularly its relationship with speech rate. It would also be beneficial to confirm that this effect can exist within speakers.

The lack of interaction between method of elicitation and metronome speed suggests that this effect can be independent of speech rate, but the connection cannot be discounted in the current study. It is still possible that individual words were uttered more quickly by the word reading participants, and that this word length was what caused the measured effect. Unfortunately, due to the choice of the carrier sentence and target words, accurately estimating the word length in the collected recordings would be very difficult –- many items end in vowels or semivowels, and establishing the boundary between them and the following /j/ of *jest* is not easy, especially in fast speech. Future research investigating this possibility should thus consider using setups and stimuli which would make word length easier to measure reliably. This should then be used to determine whether the elicitation method effect is mediated via speech rate, or if it is at least partially a separate phenomenon. If it turns out to be separate, the difference could be related to the different neural pathways engaged in each task when participants access the relevant words' phonological representations in their memory.

5 Conclusions

The behavior of Polish voiced stops mirrors that found for prevoiced stops in other languages: as the rate of speech decreases, their prevoiced portion lengthens significantly. An analogous,

VOTs. If the target VOT is too long, it becomes too difficult to sustain the glottal vibration and a partially prevoiced stop is produced.

albeit smaller effect was also found in Polish voiceless stops, which agrees with only some previous studies on short-lag stops. Those two facts are contradictory in the framework of Laryngeal Realism, thus it is crucial that the lengthening effect in short-lag stops be examined more closely. It is also recommended that we reassess whether this effect is phonological in nature, as there is a possibility it could be a phonetic, surface effect of contraction of all sounds in faster speech. Due to the difficulties of analyzing such VOT data (and production data in general) in theories presented as alternatives to Laryngeal Realism, this study suggests that we need a theory more capable of analyzing both production and perception of stop consonants. Such theory should also look at voicing beyond the parameter of VOT.

It was also found that using different elicitation methods can generate differences in VOT of Polish voiced stops. It should be investigated what other parameters measured in phonetic studies can be affected that way, and whether similar effects can be found for other languages. Another key issue that needs to be examined is whether this effect is mediated through some measure of speech rate, or if it is independent from its influence on VOT.

References

Abramson, A. S., & Whalen, D. H. (2017). Voice onset time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, *63*, 75–86. https://doi.org/10.1016/j.wocn.2017.05.002

Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, *43*, 761–770. https://doi.org/10.3758/s13428-011-0075-y

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., ... Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, *39*, 39–49. https://doi.org/10.1016/j.wocn.2010.11.001

Beckman, J., Jessen, M., & Ringen, C. (2013). Empirical evidence for laryngeal features: Aspirating vs. true voice languages. *Journal of Linguistics*, *49*(2), 259–284. https://doi.org/10.1017/s0022226712000424

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* [Computer software] (6.3.09). Retrieved April 29, 2023, from http://www.praat.org/

Cyran, E. (2014). *Between Phonology and Phonetics – Polish Voicing*. Mouton de Gruyter. https://doi.org/10.1515/9781614515135

Hall, T. A. (2001). Introduction: Phonological representations and phonetic implementation of distinctive features. In T. A. Hall (Ed.), *Distinctive Feature Theory* (pp. 1–40). Mouton de Gruyter.

Honeybone, P. (2005). Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In M. van Oostendorp & J. van de Weijer (Eds.), *The Internal Organization of Phonological Segments* (pp. 319–354). Mouton de Gruyter.

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association*, *33*(1), 103–107. https://doi.org/10.1017/S0025100303001191

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, *25(2)*, 143–168. https://doi.org/10.1006/jpho.1996.0039

Kharlamov, V. (2014). Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences. *Journal of Phonetics*, *43*, 47–56. https://doi.org/10.1016/j.wocn.2014.02.002

Lanwermeyer, M., Fanta-Jende, J., Lenz, A. N., & Korecky-Kröll, K. (2019). Competing norms of standard pronunciation. Phonetic analyses on the ‹-ig›-variation in Austria. *Dialectologia et Geolinguistica*, *27*, 143–175. https://doi.org/10.1515/dialect-2019-0008

Lee, S. S., & Berkson, K. H. (2019). Speech rate effects on VOT in a 3-category language: Evidence from Hakha Chin. *Indiana Working Papers in South Asian Languages and Cultures*, *1*(1).

Maddieson, I. (1997). Phonetic universals. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 619–639). Blackwell Publishing.

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

Malisz, Z., & Żygis, M. (2015). Voicing in Polish: Interactions with lexical stress, speech rate, focus and final boundary. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* University of Glasgow.
Paper number 963.1-5 retrieved from http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0963.pdf.

Mitterer, H. (2018). Not all geminates are created equal: Evidence from Maltese glottal consonants. *Journal of Phonetics*, *66*, 28–44. https://doi.org/10.1016/j.wocn.2017.09.003

Oh, E. (2009). Voice onset time of Korean stops as a function of speaking rate. *Phonetics and Speech Sciences*, *1*(3), 39–48.

Pająk, B., & Baković, E. (2010). Assimilation, antigemination, and contingent optionality: The phonology of monoconsonantal proclitics in Polish. *Natural Language & Linguistic Theory*, *28*(3), 643–680.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Schwartz, G., & Arndt, D. (2018). Laryngeal Realism vs. Modulation Theory – evidence from VOT discrimination in Polish. *Language Sciences*, *69*, 98–112. https://doi.org/10.1016/j.langsci.2018.07.001

Schwartz, G., Wojtkowiak, E., & Brzoza, B. (2019). Beyond VOT in the Polish laryngeal contrast. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2383–2387). Australasian Speech Science; Technology Association Inc.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 174–215.

Strycharczuk, P. (2012). *Phonetics-phonology interactions in pre-sonorant voicing* [Doctoral dissertation, University of Manchester].

Wojtkowiak, E., & Schwartz, G. (2022). The acoustics of plosives and the formation of prosodic structure in Polish. *Folia Linguistica*, *56*(2), 273–324.

Zhang, H., Carlson, M. T., & Diaz, M. T. (2019). Investigating the effects of phonological neighbours on word retrieval and phonetic variation in word naming and picture naming paradigms. *Language, Cognition and Neuroscience*, *35*(8), 980–991. https://doi.org/10.1080/23273798.2019.1686529

Appendix A: List of stimuli

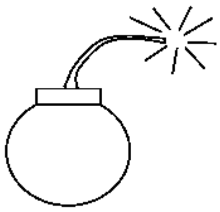
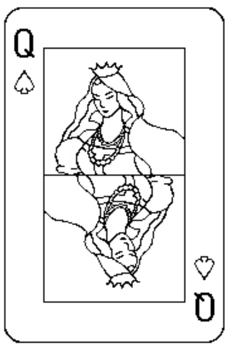
Image stimulus	Word stimulus	English translation	Stimulus type
	banan	banana	analyzed
	bomba	bomb	analyzed
	dama	queen (in card games)	analyzed
	doktor	doctor	analyzed

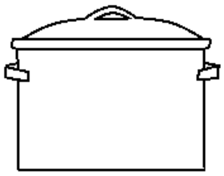
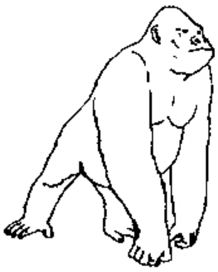
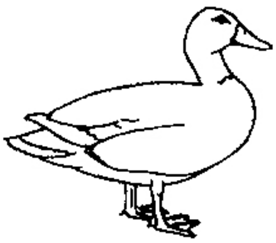
Image stimulus	Word stimulus	English translation	Stimulus type
	garnek	pot	analyzed
	goryl	gorilla	analyzed
	kaczka	duck	analyzed
	kościół	church	analyzed

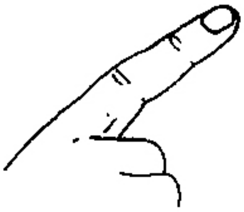
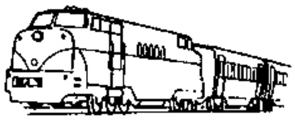
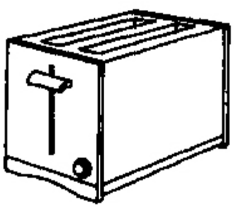
Image stimulus	Word stimulus	English translation	Stimulus type
	palec	finger	analyzed
	pociąg	train	analyzed
	taśma	tape	analyzed
	toster	toaster	analyzed

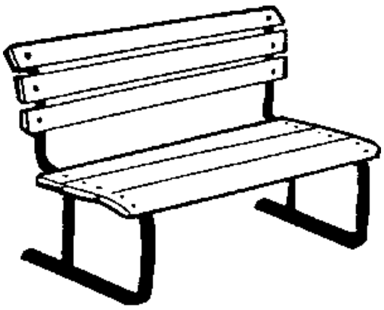
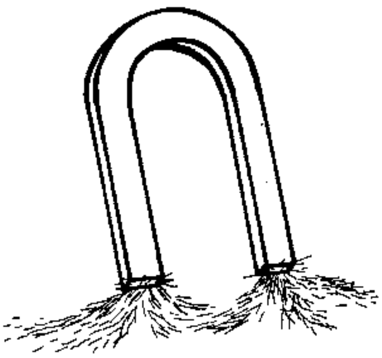
Image stimulus	Word stimulus	English translation	Stimulus type
	indyk	turkey	filler
	laska	walking cane	filler
	ławka	bench	filler
	magnes	magnet	filler

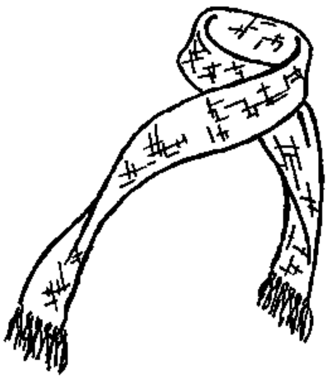
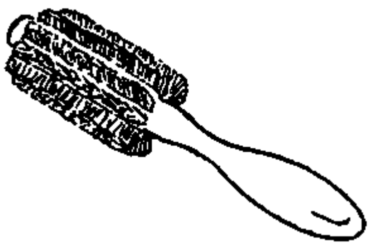
Image stimulus	Word stimulus	English translation	Stimulus type
	orzech	nut	filler
	ryba	fish	filler
	szalik	scarf	filler
	szczotka	brush	filler

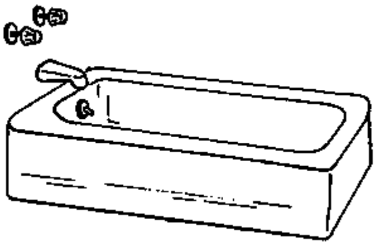
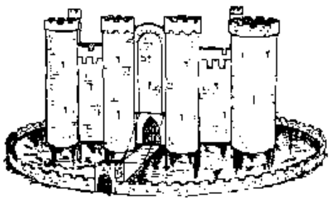
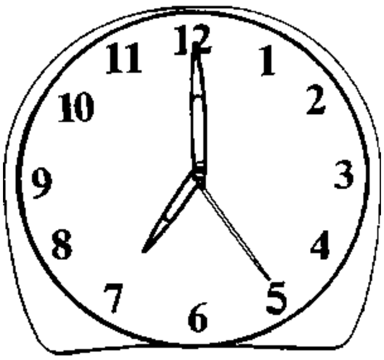
Image stimulus	Word stimulus	English translation	Stimulus type
	wanna	bathtub	filler
	wrotka	roller skate	filler
	zamek	castle	filler
	zegar	clock	filler

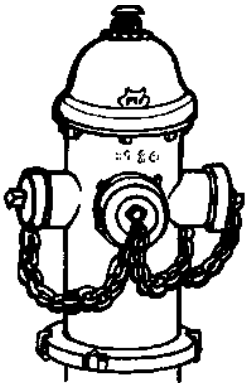
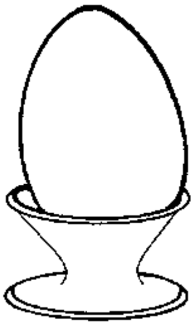
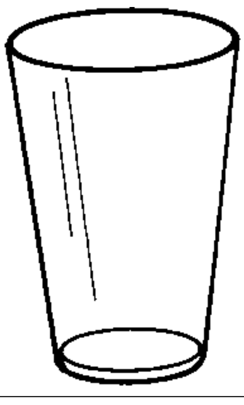
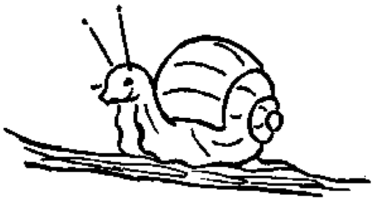
Image stimulus	Word stimulus	English translation	Stimulus type
	hydrant	fire hydrant	practice
	jajko	egg	practice
	ogień	fire	practice
	szklanka	glass (vessel)	practice

Image stimulus	Word stimulus	English translation	Stimulus type
	ślimak	snail	practice