Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Perception of Geminates in Finnish and Polish

Michael O'Dell[1], *Zofia Malisz*[2]

[1]University of Tampere, Finland
[2]Saarland University, Saarbrücken, Germany
michael.odell@uta.fi, zmalisz@coli.uni-saarland.de

Abstract

Finnish and Polish subjects judged four series of two-syllable stimuli varying in intervocalic consonant duration as short (singleton) or long (geminate). The four series differed in the durations of the flanking vowels. As predicted, perception shifted toward geminates for Finnish subjects when V_1 was lengthened, especially when V_2 was also shortened, to match normal Finnish production. The shift also occurred for Finnish subjects when V_1:V_2 ratio was kept constant. Any effect of lengthened stimuli on perceived tempo, expected to produce a geminate boundary shift in the opposite direction, was smaller than the V_1-effect. We also hypothesised that Polish subjects would show a similar pattern in perception, given that, like Finnish, Polish has slightly longer vowels before geminates. Polish listeners did present a clear sensitivity to flanking vowel durations but the observed shifts in geminate boundaries for the four stimulus series were almost exactly reversed compared to the Finnish subjects.
Index Terms: quantity, geminate, relative duration, tempo, Finnish, Polish

1. Introduction

In this study we are interested in the mechanisms of consonantal quantity perception. Traditionally quantity has been defined as duration "when it functions as an independent variable in the phonological system of a language" [1, p. 42]. However, given the multitude of factors affecting the temporal variability of speech, perception of quantity cannot be based on absolute duration alone. Consider the intervening effect of speech rate: a long consonant/vowel at a fast tempo can be even shorter than a corresponding short consonant/vowel at a slower tempo. A way to reduce the potential dependence of quantity on tempo is to suppose that language users perceive quantity based on *relative* durations (temporal relations) rather than *absolute* durations. Also, transformed or normalized relative durations (cf. [2]) can handle the fact that some segments show greater temporal variation under tempo changes than others, for instance, vowel duration varies more than consonant duration [3].

Additionally, even if tempo is controlled, quantity is not just a function of a segment's duration. In some quantity languages (e.g. Italian [4, 5]), vowels are shorter preceding geminates, whereas in others (e.g. Japanese [6]) they are longer, a difference which may be related to the particular prosodic organization (e.g. syllable vs. mora timing [7]). In any case, such dependencies can be taken to mean that quantity operates at a higher level, qualifying as a *supra*segmental property [8].

In Finnish, there are eight possible quantity patterns for a (C)-V-C-V structure [9, 10]. The Finnish quantity system is usually analyzed phonologically in terms of binary oppositions for both consonants and vowels, yielding the attested number of contrasts: $2 \times 2 \times 2 = 8$ (word initial consonants are quantity neutral). However, the realized duration of a segment is highly dependent on the quantity pattern it is part of, not just on its own phonological length. For instance, short V_1 is longer before a geminate (long C), but long V_2 is shorter [9, 10]. These dependencies mean that the eight quantity patterns are distinct not only in terms of absolute durations, but also in terms of relative durations (V_1:C:V_2 ratios), potentially facilitating discrimination by reducing dependency on context effects such as tempo. Thus a VVCCVV pattern is not simply a longer version of VCV, the ratios are different as well. This and other relations can be seen in Figure 1, which shows the eight Finnish patterns plotted on a logarithmic scale (or equivalently a two dimensional logit scale), so that all differences represent temporal ratios.

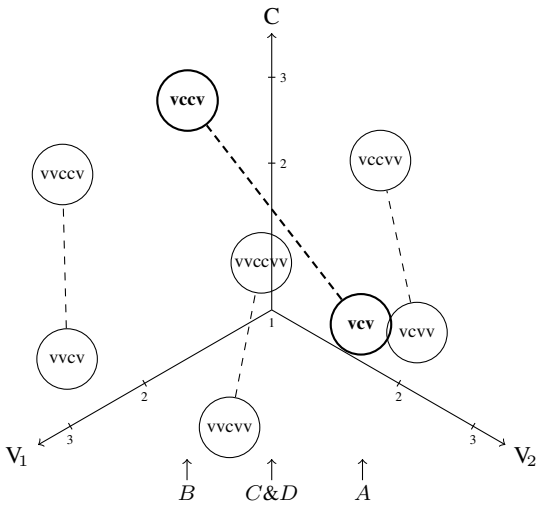

Figure 1: *Relative average durations for quantity patterns in Finnish (calculated from [9, Table 13]). C–CC pairs are joined with a dashed line. Small arrows at the bottom indicate V_1:V_2 ratios for the stimulus series A, B, C & D.*

Polish geminates provide an interesting comparison to Finnish. On the one hand, the Polish quantity system is much smaller: only consonants participate. This means, for instance, that in a V_1-C-V_2 structure there is only a two-way opposition, VCV vs. VCCV, as in the word pair *leki* vs. *lekki*, instead of the eight contrasts of Finnish. Unlike Finnish, however, some consonants exhibit a quantity contrast also in word initial position. In addition, the functional load of the quantity opposition in Polish is much smaller, since in Polish, geminates are not as ubiquitous as in Finnish. On the other hand, Polish quantity is similar to Finnish and other languages, such as Japanese, in that

 doi: 10.21437/SpeechProsody.2016-228

Polish has also been found to lengthen vowels before geminates [11, 12]. Unlike Finnish, no effect of geminates on the following vowel has been found for Polish [12]. In terms of relative durations, however, the V_1:V_2 ratio differs for singletons and geminates in Polish in much the same way as it does in Finnish, albeit not as dramatically.

Perception of quantity is also affected by factors other than the duration of the segment in question, as shown by many previous studies (e.g. [13] for Bengali, [9, 10] for Finnish). Preceding vowel duration has been reported to affect geminate perception in several languages [14], in ways matching their production patterns, e.g. geminate perception was more likely after a lengthened vowel in Japanese, but less likely in Italian.

In this paper, we concentrate on the effect of flanking vowel durations on geminate perception. Given that in Finnish a phonologically short vowel is robustly longer before geminates (by ca. 10 ms), we might expect a slightly lengthened V_1 to bias perception in favor of geminates. We call this a *V_1-effect*. Given that Polish also lengthens vowels somewhat before geminates, it is possible that a V_1-effect will be observed in Polish perception as well. Since Finnish speakers on average produce shorter vowels after geminates, it is also possible that a shorter V_2 will bias Finnish perception in favor of geminates. We call this a *V_2-effect*.

Alternatively, longer vowels (V_1 or V_2) could also signal a slower tempo, which would be expected to bias perception in favor of singletons. We call this a *tempo-effect*. For instance, one study [15] found that the perceptual geminate boundary for Japanese was fairly invariant in terms of consonant to word ratio, regardless of large vowel duration changes.

2. Methods

2.1. Stimuli

One token of the Finnish word /taakkaa/ ("burden, partitive case") recorded with Praat [16] was used as the basis for synthesizing four stimulus series with Praat's Manipulation (overlap-add technique), each of which varied [k] duration from 104 ms to 240 ms in seven steps such that the ratio of adjacent durations was approximately constant ($2^{0.2} \approx 1.1487$). The durations of [k] were 104 ms, 120 ms, 138 ms, 158 ms, 182 ms, 209 ms, 240 ms, referred to hereafter as stimulus number 1, 2, 3, 4, 5, 6 and 7. For all stimulus series pilot studies with Finnish listeners confirmed that the stimulus with the shortest [k] (104 ms) sounded unambiguously like Finnish /taka/, and the opposite endpoint with the longest [k] (240 ms) sounded unambiguously like Finnish /takka/.

The four stimulus series differed from each other in regard to the durations of the flanking vowels. Vowel durations for the first two stimulus series, A and B, were based on acoustic measurements from production studies of Finnish. Series A was synthesized using V_1 and V_2 durations appropriate for Finnish /taka/ (judging by production data from several sources, e.g. [9]): V_1 64 ms, V_2 105 ms. Series B was synthesized using V_1 and V_2 durations appropriate for Finnish /takka/ (judging by production data): V_1 73 ms, V_2 46 ms. It was expected, at least for Finnish listeners, that the perceptual boundary or crossover from /taka/ to /takka/ might be shifted in the two series, so that series A with /taka/-like vowel durations would attract more *taka* responses, and the reverse for series B.

Two additional series, here called series C and series D, were synthesized with V_1 durations identical to series A and B, but with V_2 durations equal in each case to corresponding

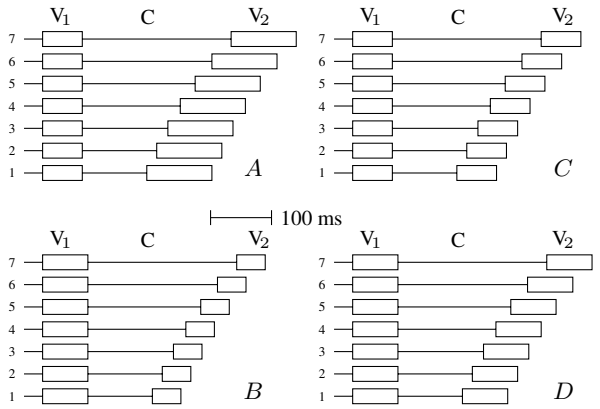

Figure 2: *Four stimulus series shown schematically.*

V_1 duration. In other words, series C had $V_1 = V_2 = 64$ ms, series D had $V_1 = V_2 = 73$ ms. The four series are shown schematically in Figure 2.

All stimuli were synthesized with a flat F0 contour at 100 Hz (approximately the value at the middle of the original token) to avoid the possibility that rate of changing pitch might cue tempo and affect quantity perception [10]. In spite of this change the stimuli sounded completely natural.

2.2. Perception test

Perception tests were executed using Praat's ExperimentMFC. Subjects pressed one of two keys, **D** for *taka* and **K** for *takka*. Each stimulus was presented ten times, giving a total of $7 \times 4 \times 10 = 280$ responses for each subject. All stimuli were presented in one session with independent randomization with no stimulus presented twice sequentially. Subjects were given the chance and encouraged to take a break after every 70th stimulus. The test was administered to ten adult native speakers of Finnish, eight females and two males, and ten adult native speakers of Polish, five females and five males.

2.3. Statistical analysis

A Bayesian hierarchical logistic regression model [17] was fit to the perception data using JAGS [18]. For each LANGUAGE (Finnish, Polish) an intercept parameter (50 % crossover, indicating the boundary location between /taka/ and /takka/ perception) was allowed to vary with covariates SUBJECT, SERIES (A, B, C or D), and SERIES×SUBJECT interaction. A slope parameter, indicating how rapidly perception changes from /taka/ to /takka/, was allowed to vary by SUBJECT. In addition, two parameters, allowed to vary by SUBJECT, were included for floor and ceiling effects, i.e. the minimum and maximum possible probability of geminate perception.

3. Results

The posterior distribution for the standard deviation of effects on the intercept (geminate boundary) is summarized in Figure 3 using so called caterpillar plots for credible intervals (CI; thin line indicates 95 % CI, thick line 50 % CI, dot indicates posterior median). The x-axis is scaled in terms of steps in the stimulus series. Figure 3 shows a clear effect of LANGUAGE. Both SUBJECT and SERIES clearly made a difference in the intercept for both languages. The size of both SUBJECT and SERIES

differences is very similar for the two languages. If there are SERIES×SUBJECT interactions, they are relatively small.

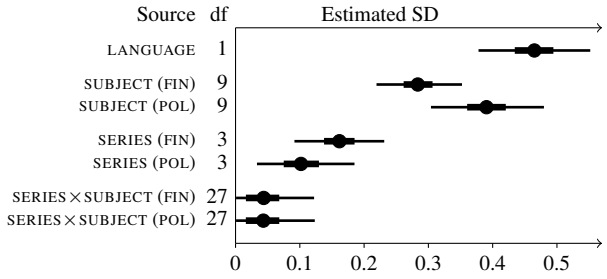

Figure 3: *Estimated SD of effects on geminate boundary.*

3.1. General differences between Finnish and Polish

We first present the general differences between Finnish and Polish subjects, and then move on to the differences associated with the four stimulus series.

Intercept: In general, that is, ignoring SERIES and SUBJECT related effects, the 50 % intercept, i.e. geminate boundary was shifted towards longer [k] for the Polish subjects compared to the Finnish subjects. The posterior medians for Finnish and Polish were 3.78 and 4.43, respectably, corresponding to 153.5 and 168.1 ms, 95 % CI were (3.71, 3.84) and (4.33, 4.54), corresponding to (152.1 ms, 154.9 ms) for Finnish and (165.8 ms, 170.5 ms) for Polish.

Slope: In general the slope of the logistic curve was slightly steeper for Finnish than for Polish. The posterior medians for Finnish and Polish were 2.99 and 2.16, respectably, with 95 % CI of (2.66, 3.37) for Finnish and (1.87, 2.52) for Polish. Thus in general the quantity boundary was sharper for the Finnish subjects, although there was also significant variation among subjects in both groups.

Floor and ceiling effects: Lower and upper limits for probability of geminate perception were practically 0 % and 100 % for the Finnish subjects (95 % CIs (0.0 %, 0.4 %) and (99.5 %, 100.0 %)). For the Polish subjects, on the other hand, there was some slight residual uncertainty even at the endpoints of the stimulus series: general (ignoring SUBJECT differences) median posterior values for floor and ceiling probabilities were 2.6 % and 98.4 % (95 % CIs (0.3 %, 14.1 %) and (91.4 %, 99.8 %)).

Both the differences in slope and the differences in lower and upper limits reflect the fact that the quantity opposition is not as heavily used in Polish compared to Finnish. All these general findings are summarized in Figure 4, which shows the logistic regression curves for geminate perception in Finnish and Polish using posterior median values for intercept, slope, floor and ceiling.

3.2. Effect of vowel duration on geminate boundary

Figure 5 shows the effects of individual stimulus series on intercept in general, without SUBJECT effects. Again, the x-axis is scaled in terms of steps in the stimulus series. For Finnish, the posterior probability that the geminate boundary for series A is shifted to the right compared to series $B$is $\Pr(A > B) > 0.9999$. The corresponding posterior probability that the boundary for series C is shifted to right compared to series $D$is $\Pr(C > D) = 0.9434$. Interestingly, for Polish, the influence

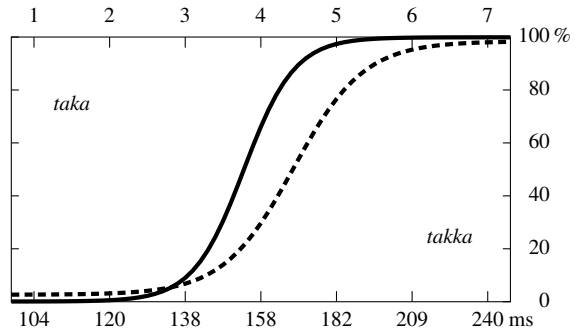

Figure 4: *General (i.e. no* SERIES *or* SUBJECT *effects) geminate perception by* [k] *duration for Finnish (solid) and Polish (dashed). Parameters have posterior median values.*

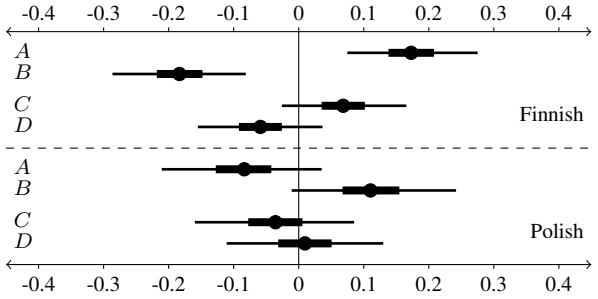

Figure 5: *Estimated effects of stimulus series on intercept (CI) for Finnish subjects (top) and Polish subjects (bottom).*

of A vs. B and C vs. D is reversed. The probabilities for Polish without SUBJECT effects equal $\Pr(B > A) = 0.9718$ and $\Pr(D > C) = 0.6715$.

Since the SERIES×SUBJECT effects are very small, these results are also representative of individual subjects. Table 1 displays the posterior probability by subject and shows (i) that the geminate boundary for series A is shifted to the right compared to series B, and (ii) that the geminate boundary for series C is shifted to the right compared to series D. Note that for all Finnish subjects, the probability $\Pr(A > B)$ is very high (always > 0.98), and the probability $\Pr(C > D)$ is fairly high (always > 0.75, often > 0.90). For the Polish subjects, the probability $\Pr(A > B)$ is fairly low (always < 0.10), which

Table 1: *Probability of the intercept difference by subject.*

	Finnish subjects		Polish subjects	
	$A > B$	$C > D$	$A > B$	$C > D$
1	0.9987	0.9390	0.0704	0.3408
2	0.9970	0.8249	0.0794	0.4186
3	0.9892	0.9159	0.0563	0.3645
4	0.9998	0.8726	0.0395	0.3119
5	0.9831	0.9047	0.0687	0.4009
6	0.9982	0.8715	0.0832	0.2644
7	0.9945	0.7756	0.0635	0.3928
8	0.9827	0.8949	0.0541	0.3299
9	0.9985	0.9182	0.0874	0.4092
10	0.9966	0.8026	0.0812	0.3688

means that the opposite probability $\Pr(B > A)$ is fairly high. The posterior probability $\Pr(C > D)$ is always < 0.45, which means that $\Pr(C < D)$ is always > 0.55.

4. Discussion

4.1. Finnish

The geminate boundary was clearly shifted toward shorter [k] durations in series B compared to series A for our Finnish subjects, a result consistent with a V_1-effect, since on average Finnish speakers produce slightly longer phonologically short vowels before geminates. This result is also consistent with a V_2-effect, with the shorter V_2 of series B responsible for increased geminate perception, since Finnish speakers on average produce *shorter* vowels *after* geminates. Of course it is entirely possible that both effects are responsible for the patterns observed.

Yet a third explanation is available, the tempo-effect, based on the fact that quantity distinctions cannot be independent of perceived speech rate. It is possible that because the total vowel duration for series A (169 ms) is greater than for series B (119 ms), stimuli in series A, being longer in total duration than the corresponding stimuli in series B, sound slightly slower and therefore slightly longer durations of [k] are required to produce a geminate perception.

Series C and D were designed to shed more light on this issue. In these two series the V_1:V_2 ratio was kept constant (at one). If quantity perception is based solely on relative durations, we would expect the boundary to be shifted by approximately one stimulus series step: the series were designed so that a stimulus n in series C has approximately the same relative durations as stimulus $n + 1$ in series D. If the geminate boundary is shifted for C vs. D, the direction of the shift should also tell us something about the perceptual importance of V_1 duration vs. V_2 duration, as well as the strength of a tempo-effect.

Although the difference was smaller than for series A vs. B, the geminate boundary was shifted toward *shorter* [k] durations in series D compared to series C. In other words, the probability of geminate perception was increased for series D, which had longer V_1 as well as longer V_2. The empirical difference is thus the opposite of that predicted by the relative duration hypothesis. It is, however, compatible with the V_1-effect (see above), but contrary to the V_2-effect and contrary to the tempo-effect as well. This would seem to indicate that V_1 duration had more influence on geminate perception than either V_2 duration or a speech tempo effect. Obviously increasing the duration of both V_1 and V_2 to a still greater extent should eventually cause the word to sound slow enough to push the geminate boundary back again towards longer durations.

4.2. Polish

Based on the fact that Polish, like Finnish, has slightly longer vowels before geminates, our hypothesis was that manipulating vowel durations might have similar effects on perception in the two languages. Looking at Figure 3, it is clear that in both languages varying the duration of the flanking vowels did have a definite effect on the perceived geminate boundary. However, looking at Figure 5 and Table 1, it is equally clear that the effect was very different for our Polish subjects compared to our Finnish subjects.

The V_1-effect applied to Polish predicts that the geminate boundary should shift toward shorter [k] in the series with longer V_1, given that in production studies Polish vowels have been found to be longer before geminates. That is, we should observe more geminate perceptions in B (compared to A) and D (compared to C). Instead, for our Polish subjects, the boundary is shifted toward *longer* [k] durations for both B and D.

The V_2-effect hypothesized for Finnish makes no prediction for Polish, given that in production studies geminates had no effect on V_2 in Polish. The tempo hypothesis considered above, that total stimulus duration affects perceived tempo, and that geminate perception in turn depends on tempo, would predict that the stimulus series should be ordered as follows (from short to long geminate boundary): B, C, D, A. This prediction is consistent with the order $C < D$ for our Polish subjects, but the order $A < B$, with an even larger geminate boundary shift, is problematic. If a tempo-effect is at work here, it must be overridden by an even stronger effect related to the vowel durations in series A and B. An explanation based on relative durations is also consistent with $C < D$ (although the size of the empirical shift for Polish is much too small).

A possible reason for the observed shift is that it may be due to the very short duration (46 ms) of V_2 in series B, since such short vowels are not expected in Polish, especially if words are pronounced in isolation, but the possible mechanics of such an effect are not clear. In any case, it is perhaps relevant that all stimuli were derived from an original Finnish language production (although all durations were greatly modified), and also that stimulus series A and B in particular, were designed with Finnish durational patterns in mind. This fact may also partially explain the extra uncertainty of the Polish subjects. In the future, we plan to reverse the experimental setup presenting stimuli based on Polish production to both Finnish and Polish listeners.

5. Conclusions

The small production differences observed in the durations of the flanking vowels for singleton vs. geminate consonants in Finnish were matched in perception. When V_1 was slightly lengthened and V_2 was simultaneously shortened to match typical Finnish CVCCV production compared to CVCV, the perceptual boundary shifted toward geminates for our Finnish listeners. Somewhat unexpectedly, our Finnish subjects also heard more geminates when V_1 duration was increased and V_1:V_2 ratio was kept constant (however the shift was smaller). This implies that quantity perception cannot be based solely on *relative* durations, even allowing different segments to have different weights, since increasing vowel duration *decreases* the relative duration of the consonant. Apparently the effect of the lengthened V_1 on perceived quantity pattern overrode any other effects. Because all stimuli were presented in one batch, subjects may have been predisposed to assume a *constant* tempo. Consequently, the vowel differences in the stimulus series were not large enough to overturn that assumption.

We hypothesized that Polish subjects might show a similar pattern in perception, given that, like Finnish, Polish also has slightly longer vowels before geminates in production. While Polish listeners did show a clear sensitivity to flanking vowel duration, the observed shifts in geminate boundaries were almost exactly reversed compared to the Finnish subjects. While the reasons for this reversal are not clear and necessitate additional research, this result does highlight the fact that languages can be quite different with regard to details of quantity, in spite of superficial similarities.

6. References

[1] I. Lehiste, *Suprasegmentals*. Cambridge MA: Massachusetts Institute of Technology Press, 1970.

[2] G. Chung, "Hierarchical duration modelling for a speech recognition system," Master's thesis, M.I.T., 1997. [Online]. Available: http://dspace.mit.edu/bitstream/handle/1721.1/42666/37658973-MIT.pdf

[3] V. A. Kozhevnikov and L. A. Chistovich, *Speech, Articulation, and Perception*. Washington, DC: Joint Publications Research Service, 1965, translation of Речь. Артикуляция и восприятия. Москва-Ленинград, «Наука».

[4] A. Esposito and M. G. Di Benedetto, "Acoustical and perceptual study of gemination in Italian stops," *Journal of the Acoustical Society of America*, vol. 106, no. 4, pp. 2051–2062, 1999.

[5] C. Zmarich, B. Gili Fivela, P. Perrier, C. Savariaux, and G. Tisato, "Speech timing organization for the phonological length contrast in Italian consonants," in *Proceedings of Interspeech 2011*, Florence, Italy, 2011, pp. 401–404.

[6] K. Idemaru and S. G. Guion, "Acoustic covariants of length contrast in Japanese stops," *Journal of the International Phonetic Association*, vol. 38, no. 2, pp. 167–186, 2008.

[7] W. H. Ham, *Phonetic and Phonological Aspects of Geminate Timing*. New York: Routledge, 2001.

[8] J. Local and A. Simpson, "Phonetic implementation of geminates in Malayalam nouns," in *Proceedings of the XIVth International Congress of Phonetic Sciences*, J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey, Eds. University of California, Berkeley, 1999, pp. 595–598.

[9] J. Lehtonen, *Aspects of Quantity in Standard Finnish*, ser. Studia Philologica Jyväskyläensia. University of Jyväskylä, 1970, no. VI.

[10] M. O'Dell, *Intrinsic Timing and Quantity in Finnish*, ser. Acta Universitatis Tamperensis 979. Tampere University Press, 2003. [Online]. Available: http://acta.uta.fi/teos.phtml?9540

[11] Z. Malisz, "Vowel duration in pre-geminate contexts in Polish," in *Interspeech, Brighton, UK, 2009*, 2009, pp. 1547–1550.

[12] ——, "Speech rhythm variability in Polish and English: A study of interaction between rhythmic levels," Ph.D. dissertation, Faculty of English, Adam Mickiewicz University, 2013.

[13] J. Hankamer, A. Lahiri, and J. Koreman, "Perception of consonant length: Voiceless stops in Turkish and Bengali," *Journal of Phonetics*, vol. 17, pp. 283–298, 1989.

[14] J. Kingston, S. Kawahara, D. Chambless, D. Mash, and E. Brenner-Alsop, "Contextual effects on the perception of duration," *Journal of Phonetics*, vol. 37, no. 3, pp. 297–320, 2009.

[15] S. Amano, R. Mugitani, and T. Kobayashi, "Perceptual boundary between a single and a geminate stop in Japanese," in *Proceedings of the 17th International Congress of Phonetic Sciences*, W.-S. Lee and E. Zee, Eds. City University of Hong Kong, 2011, pp. 733–736.

[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 6.0) [computer program]," 2015. [Online]. Available: http://www.praat.org/

[17] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2007.

[18] M. Plummer, "JAGS: Just Another Gibbs Sampler (version 4.1.0) [computer program]," 2016. [Online]. Available: https://sourceforge.net/projects/mcmc-jags/