ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2008
Campinas, Brazil
May 6-9, 2008

# A Preliminary Study of Temporal Adaptation in Polish VC Groups

*Zofia Malisz and Katarzyna Klessa*

School of English, Adam Mickiewicz University, Poznań
Institute of Linguistics, Adam Mickiewicz University, Poznań
zmalisz@ifa.amu.edu.pl, klessa@amu.edu.pl

## Abstract

The study presents experimental data on Polish vowel durations in consonantal contexts gathered to test prosodic hypotheses. An attempt is made to verify the significance of the process of balancing V-to-V durations in a dynamical model of speech rhythm applied to Polish. We report on the results of a controlled experiment followed by a query of a corpus of running Polish speech, both investigating the influence of selected Polish consonants on preceding vowel duration.

## 1.Introduction

The voicing effect, i.e. the lengthening of a vowel preceding a voiced consonant, is a very well documented phenomenon in many languages [4] especially in the Germanic family [18, 19, 21]. The adaptation of duration between vowels and consonants seems to be not only a cue to the following consonant voicing but also a direct compensation of length between segments, a phonetic representation of temporal planning. This prosodic aspect of the effect was signaled to be relevant for the balance of syllable duration [10, 14] and to have consequences for dynamical rhythm models such as [2].

The speech rhythm model in [1, 2] is specified by the coupling of syllable-sized and phrase stress oscillator, where the first provides regularity and the other structure. The syllabic oscillator is implemented by V-to-V units. Given the hypothesis that there exists a tendency to regularize the recurrence of vowel onsets, duration compensation inside the V-to-V unit would support periodic oscillation. Intervals of voiceless-longer/voiced-shorter consonants within a VCV frame act as perturbing factors. In Barbosa's [2] study of Brazilian Portuguese, interpreted for the purposes of the rhythm model, the relative imbalance in duration between segments in the voicing effect environments, tended to trigger temporal adaptation on the part of the vowel. Moreover, when combinations of voicing, manner and place in the following context were selected, such as /CVrV/ vs. /CVSV/, reflecting significant durational differences between consonants, vowel adaptation occurred as well [see also 5].

According to Keating's work on consonant voicing categorisation [9], the phonemic voicing effect does not occur before Polish stops (see also [10] for Czech and [16] for Arabic) even though the difference of length between the consonants is significant. The presence of the effect, on the other hand, was reported by [22] for vowels in isolated nonsense words, as well as by [8]. Recent Polish corpus data results [3, 17] confirm only some contextual voicing effects on Polish mean vowel duration. We are not aware of any previous studies investigating mutual contextual duration effects of consonants and vowels in Polish. A recent study [15] on Czech VC and CV sequences reports some temporal adaptation inside the Czech VC.

In our previous studies on Polish segmental duration [3, 13] for speech synthesis purposes we tested over 50 features both from the segmental and suprasegmental levels with CART algorithm to verify their correlation with phone duration. The influence of the right context was rated high in the rankings obtained from three large corpora for two feature vectors. The identity of the sound directly following the sound in question appeared to be one of the two most important features within the feature vector, and the manner of articulation was one of the first ten most important features. It is though still necessary to remember that it is only one out of over fifty other features whose complex interactions should not be neglected while analyzing particular aspects of speech timing.

## 2.Data and Methods

Two datasets were used to verify our hypothesis: a set of phrases with target items in controlled positions and a corpus of read speech data. In both sets of data we were interested in durational compensation effects or the lack thereof in VC groups within V-to-V units. All analyzed data were labeled according to prosodic annotation rules (cf. 2.3).

### 2.1. Controlled experiment

Four speakers (2 male, 2 female) of standard Polish were asked to repeat stimuli around 20 times (around thousand tokens were recorded and annotated). In this part of the study we investigated the influence of Polish alveolar, postalveolar and palatal fricatives /s, z, S, Z, s,' z'/ on preceding vowel duration. This particular set was chosen because the fricatives are known to differ more than stops in their inherent durations within the voicing contrast [12]. Only medial VC units were considered since Polish consonants do not contrast in voicing in final position: all voiced obstruents are devoiced word finally. In the controlled experiment we included contexts with stops /p, b, t, d/ as well, for comparison with Keating's original work, which investigated /raCa/ type of stimuli with stops contrasting in voicing.

We used target words of the form *kaCa* and *raCa* where C was one of the six fricatives or four stops under study. The main stimuli were presented on randomized strips of paper in the following form:
„*To nie jest kasa* _____, *to kasa* _____"
*"To nie jest kaza* _____, *to kaza* _____"

The words to be filled into the gaps were given on separate strips of paper: *dobra, tania* ("good", "cheap") for the first condition containing meaningful target words (e.g.: "kasa", "rasa", "Kasia") and *lośna, dąpna* (no meaning) for the second condition containing nonsense target words ("kaza"). These two pairs of additional stimuli functioned as masking words. In a few preparatory runs the speakers did not have problems with incorporating the nonsense elements into

a meaningful carrier sentence, producing fluent utterances.

The stimulus design placed target words in a carrier sentence in order to elicit a more natural speaking style as opposed to a word list used in [9]. The main task was constructed so that the target word would not attract prominent focus, assigned to other potential locations in the frame. The "new information" status and position in the phrase attracted phrasal prominence to the masking words rather than to the target word. This way we attempted to control emphasis pragmatically. Gibbon et al. [7] suggest that the lack of contrastive length in Polish might provide an extra degree of freedom for use in emphasis; emphasis by lengthening is liberally used in Polish.

The experiment was controlled for the following factors that influence segment duration: vowel identity, syllable shape and count, position in the phrase, lexical stress, emphasis. It was not directly controlled for rate, however we further present ratio measurements which normalize the possible rate effects.

### 2.2. Database query

The corpus consisted of speech data provided by forty speakers (20 male and 20 female, aged 20-25) reading a 25-sentence script of an informal interview with a politician.
We searched the corpus for tokens that contain VC groups containing /a/ and either a very short consonant such as /r/ or a very long one, /S/, similarly to Barbosa [2] who investigated the temporal relations within the same CV sequences in a controlled experiment (/a S/, /a r/). For the present data these types of tokens were found in four realizations by each of the forty speakers, namely in the words: /naSe/ /vaSe/ (adjectives) produced in phrase-medial position, each of them as an attribute of the word /mjasto/, and in the word /staram/ (a verb) produced twice by each speaker: phrase-initially and phrase-medially (or phrase-finally, depending on individual phrasing techniques applied by speakers).

It might have been anticipated that for the corpus data the common influence of multiple modifying factors would be inevitably stronger than for data in controlled experiments. The individual realizations of phrase accents, emphasis and pauses were carefully examined perceptually as well as on the basis of spectrogram inspections in order to obtain a possibly consistent dataset. Target words produced with a pitch or emphatic accent or with a pause in the right context were excluded from further analyses.

### 2.3. Prosodic annotation

Oral constriction criteria were applied as guidelines for prosodic annotation as described in [24]. We found that such criteria correspond better to our objective of investigating temporal relations as they illustrate the motor tasks that a speaker must dynamically tackle in the production of VC groups. The standards are different from annotation based on acoustic cues, e.g., in the following cases:

- silent transition interval following voiceless fricative noise is included in the following vowel interval.
- the silent interval occurring after a vowel before a voiceless fricative noise sets in is counted as belonging to the vocalic interval.

It is necessary to add, as we had used two types of target words, one beginning with /r/ and the other with /k/, that the annotation standard sets the final boundary for /k/ at the consonant release rather than at the beginning of voicing. Any short transition period, not more than 10 ms on average though, was included in the following vocalic interval. The effect of this annotation strategy is systematic and despite the /k/ annotation induced lengthening, the first vowels in /kaCa/ were consistently shorter in all speakers and right contexts than in /raCa/.

For the controlled speech data (cf. 2.1) the signal was segmented manually using Praat speech analysis software and annotation tools. The corpus data were initially labeled automatically with SALIAN [23] and then manually adjusted to the abovementioned rules of prosodic annotation.

## 3.Results

Fig. 1 presents the consonant-consonant and vowel-vowel duration ratios for the kaCa stimuli in the controlled experiment. It shows that on average, all the studied voiced fricatives were shorter than the voiceless ones across speakers and places of articulation by approx. 30%.

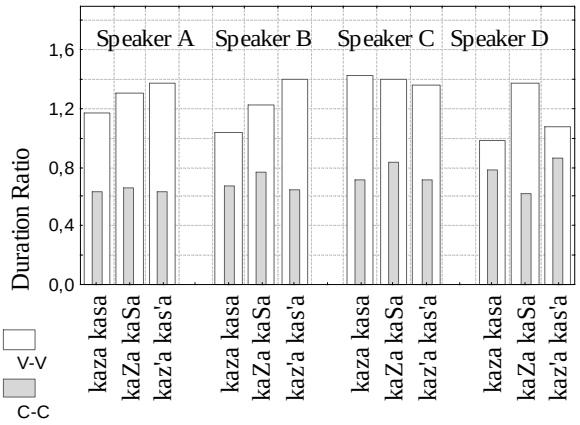


*Figure 1: Mean duration ratio for vowel pairs - /a/ preceding /z Z z'/ to /a/ preceding /s S s'/ respectively; and for consonant pairs /s z/ /S Z/ /s'z'/ in kaCa type of words.*

This is inversely proportional to the duration of the preceding vowels, which lengthen by 30%. However, the V/V ratio in the pre-voiced/pre-voiceless context respectively, displays some speaker dependent behaviour.

Fig. 2, for target words of the form /kaCa/ (/kasa/, /kaza/, /kaSa/ etc.), sheds some light on the matter in terms of statistical significance. All speakers produce significant differences in the context of /S, Z/ and /s', z'/. Speaker B and D did not produce a significant difference in the context of /s,z/. Speaker D's behaviour is quite consistent: where the consonant ratio comes closer to one, i.e. the homorganic fricatives are more equal in duration, the vowel ratio is lower. It appears as if the speaker aims at producing balanced vocalic and consonantal intervals.

In fact speakers A, C and D and their individual timing strategies turned out to be quite predictable in both fricative tasks. Their sentence realizations were more uniform in terms of pitch changes and phrasing. Speaker B varied her productions more from sentence to sentence, in a less repetitive, more "illustrative" style, this variability can be seen in the graphs.

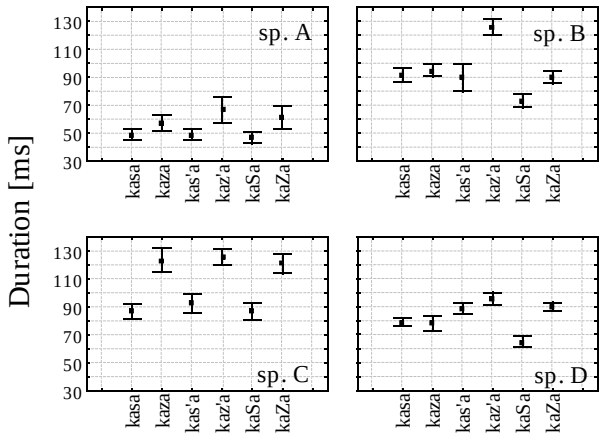


*Figure 2: Mean durations (points) with confidence intervals (lines) for /a/ preceding /s z S Z s' z' / in kaCa type of words.*

The results of our stop voicing effect study are inconclusive. For the same 4 speakers we obtained varying measures of significance. It corresponds to conflicting reports in the matter. [2] reports on the basis of similar Brazilian Portuguese data that stops do not always show significant differences for the voicing effect, which is agreed to exist in Brazilian Portuguese. Table 1 contains the means and standard deviation values, with statistically significant items in bold, for all /raCa/ words.

The /raCa/ fricative set, in spite of some missing data and higher standard deviations, possibly due to difficulties with segmenting /r/, reflects the trend found in the /kaCa/ words.

Table 1. *Mean duration values and standard deviations for the vowel /a/ depending on the following consonant context in raCa type of words. Values which produced statistically significant differences are in bold.*

| *Left Context* | *r* | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *Right Context* | *fricatives* | | | | | | *Stops* | | | |
| | *s* | *z* | *S* | *Z* | *s'* | *z'* | *p* | *b* | *t* | *d* |
| *Mean subject A* | **60** | **85** | **76** | **86** | 79 | n/a | 75 | 71 | 63 | 63 |
| *St.dev subject A* | **13** | **17** | **19** | **24** | 18 | n/a | 11 | 12 | 8 | 8 |
| *Mean subject B* | 111 | n/a | **104** | **91** | **127** | **157** | **94** | **107** | **101** | **120** |
| *Stdev subject B* | 21 | n/a | **12** | **8** | **12** | **19** | 17 | 17 | 26 | 7 |
| *Mean subject C* | **104** | **138** | **123** | **142** | 118 | 124 | 103 | 99 | 96 | 120 |
| *St.dev. subject C* | **12** | **16** | **12** | **19** | 15 | 16 | 15 | 12 | 9 | 12 |
| *Mean subject D* | 72 | 74 | 71 | 77 | 75 | 80 | **72** | **56** | 75 | 71 |
| *St.dev. subject D* | 10 | 9 | 10 | 11 | 10 | 10 | 11 | 7 | 8 | 7 |

As far as the differences between consonants are concerned, all fricative pairs differed significantly in length, confirming our assumptions that fricative duration across the voicing distinction is clear cut and so may trigger vowel length adaptation. Stop durations produced one insignificant result. In speaker D's /rapa/-/raba/ pair, the consonants were equally long, the difference in vowel duration was significant, however, in the opposite direction than expected: the vowel in front of /b/ was shorter but not as short as the one in front of /p/. The case appears to be an outlier.

We also compared the mean duration of /p/ to /S/ and /s'/, assuming these are the longest consonants within the two manners. The mean values of /p/ among speakers were lower than the approximately equal /S/ and /s'/ values by 2 to 10ms only.

### 3.1. Corpus query response

For the corpus data, the effect of compensation depending on the following long versus short consonant appeared to be less significant than for the controlled speech data, however the tendency was not contradicted.

The V-V ratio after excluding items produced with a pitch accent on /a/ or with a pause following the target word was as follows: 1,39 (for phrase-medial position of the source word for /ar/) and 1,03 (phrase-initial position of the source word for /ar/). The ratios of consonant durations were: 0,25 and 0,33 respectively.

Table 2. *Mean duration values and standard deviation for the vowel /a/ depending on the following consonant context for four words produced by 40 speakers (PI-Phrase Initial, PM-Phrase Medial)*

| Source word | V duration | Std. dev | C duration | Std. dev |
|---|---|---|---|---|
| /vaSe/ | 59,87 | 10,6 | 99,5 | 12,6 |
| /naSe/ | 58,23 | 9,9 | 104,34 | 10,9 |
| /staram/ PI | 60,98 | 11 | 25,95 | 4,1 |
| /staram/ PM | 82 | 12,5 | 33,89 | 6,4 |

Table 2 presents mean duration values for vowels and consonants in VC groups realized by forty speakers. The first two rows give mean durations for the /aS/ sequences from the words /vaSe/ /naSe/ that were compared to durations from /ar/ groups in the word /staram/. The latter was produced in two positions by each speaker (initial or medial), and, as it is shown in the table, the difference between the two types of realizations appeared to be the prevalent one.

## 4.Discussion

Keating [9] reported some closure duration overlap between homorganic stops in Polish, suggesting they do not always clearly fall into two groups. Due to such past results we expected variability in the stop voicing effect interpreted in duration compensation terms, at least for some speakers. The stop effect did produce variable results. However, all but one stop duration difference proved to be significant in our study. Moreover, /p/, /S/ and /s'/ were similar in duration, but only the fricatives caused vocalic temporal adaptation. We need to conclude that manner of articulation effects may override the voicing effect before stops and support it before sibilant fricatives we studied here, with the possible exception of /s/ and /z/. This possibility needs to be looked into in the future, if

a compensatory interpretation of the voicing effect is to be found valid for Polish. Indeed, an extensive corpus study in [12] demonstrated that vowels are longer if the degree of stricture of the following consonant is lower. This finding could help explain the general damping of compensatory durations in the stop effect environments. Further study is needed as to the exact relationship between stop duration and vowel duration in Polish.

In general, our results suggest segment timing in voicing effect environments in Polish is to some extent speaker dependent [see also 12], the language does not dictate phonological rules in this domain. Some factors that influence the magnitude of phonemic voicing effect are absent in Polish: there is no contrastive vowel length, final consonant voicing contrast is not neutralized word medially. Some authors have suggested that weakening or absence of the effect may have sources on the rhythmical level [19] depending on the rhythm type. There is no agreement as to whether Polish is syllable- or stress-timed although recent work [7] suggests that Polish phonological syllables are more uniform in length than V-to-V stretches, what in turn implies some segment duration adaptation within the canonical syllable. Our results for VC groups involving fricatives however show a tendency to balance the V-to-V segment timing relations, as predicted by the model in [2]. As for stops, as discussed above, the interaction with manner of articulation needs to be explained.

Due to insufficient control of factor interaction in the tokens extracted from our running speech corpus, we see the need of constructing specifically designed corpora for isolating and analyzing temporal details of the voicing effect type. To fully validate the duration compensation hypothesis different combinations of segment durations within VC groups need to be tested.

## 5.Conclusions

The results of the experimental case study are encouraging. More elaborate investigation of running speech with a greater number of speakers would be necessary along with careful text input design to answer both the requirements of corpus size and sufficient control of the analyzed relations.

**Acknowledgements**

The running speech dataset used in the experiment is part of a speech synthesis and recognition corpora created within a project supported by The Polish Scientific Comittee (Project ID: R00 035 02).

## 6.References

1] Barbosa, P.A., 2002. Explaining Cross-Linguistic Rhythmic Variability via a Coupled-Oscillator Model of Rhythm Production. In: Proceedings of Speech Prosody 2002, Aix-en-Provence. p. 163-166.

2] Barbosa, P.A., 2006. *Incursões em torno do ritmo da fala*. Campinas: Pontes.

3] Breuer S., Francuzik K. (Klessa, K.), Demenko G. 2006 Analysis of Polish Segmental Duration with CART, Proceedings of Speech Prosody 2006, 4, Dresden.

4] Chen M. 1970. Vowel length variation as a function of the voicing of consonant environment. Phonetica 22, 129-159.

5] Crystal, T.H.; House, A.S., 1988. Segmental duration in connected speech signals: current results. Journal of the Acoustical Society of America 83.4, 1553-1573.

6] Cummins F. Port R, 1998. Rhythmic constraints on English speech rhythm. Journal of Phonetics 26(2): 45-171.

7] Gibbon, D., Bachan, J., Demenko, G. 2007. Syllable timing patterns in Polish: results from annotation mining. [in:] Proceedings of Interspeech Conference.

8] Imiołczyk, J.; Nowak, I.; Demenko, G., 1994. High intelligibility text-to-speech synthesis for Polish. Archieves of Acoustics 19 (2) 161-172.

9] Keating, P., 1979. A phonetic study of a voicing contrast in Polish. Unpublished doctoral dissertation. Brown University.

10] Keating P. 1985. Universal phonetics and the organization of grammars, Phonetic Linguistics, ed. V. Fromkin, Academic Press, 115-132.

11] Klatt, D., 1973. Interaction between factors that influence vowel duration, JASA 54, 1102-1104.

12] Klessa K. 2006. Modelowanie iloczasu głoskowego na potrzeby syntezy mowy polskiej. (Modelling segmental duration for speech synthesis purposes). Doctoral dissertation. Adam Mickiewicz University, Poznań, Poland.

13] Klessa, K., Szymański, M., Breuer, S., Demenko, G., 2007. Optimization of Polish Segmental Duration Prediction with CART., SSW6, Bonn.

14] Lehiste, I., 1977. Isochrony reconsidered. Journal of Phonetics 5.3. 253-263.

15] Machac, P.; Skarnitzl, R. 2007. Temporal compensation in Czech? Proceedings of the 16th ICPhS, Saarbruecken.

16] Mitleb, F. M., 1984. Voicing effect on vowel duration is not an absolute universal. Journal of Phonetics, v.12, 23-27.

17] Nowak P. 2006. Vowel reduction in Polish. Doctoral dissertation. University of California, Berkeley.

18] Port R. 1981. Linguistic timing factors in combination JASA 69(1), 262-174.

19] Port, R. F. ; Dalby, J., 1982. C/V ratio as a cue for voicing in English. Perception & Psychophysics 2.141–52.

20] Port, R.; Al-Ani, S.; Maeda S. 1980 Temporal compensation and universal phonetics. Phonetica, 37: 235-252.

21] Port, R. F. ; O'Dell, M.,1985. Neutralization of syllable-final voicing in German. Journal of Phonetics 13, 455–71.

22] Richter, L., 1973. The duration of Polish vowels. Speech Analysis and Synthesis, 3/1973, 87-115., Warszawa.

23] Szymański M. and Grocholewski S., Transcription-based automatic segmentation of speech. [in:] Proceedings of 2nd Language & Technology Conference, Poznań, 2005,pp. 11-15.

24] Turk A., Nakai S., Sugahara M. 2006 Acoustic Segment durations in prosodic research: a practical guide. Methods in Empirical Prosody Research, Berlin: Mouton de Gruyter.