SEGMENTAL CONTEXT EFFECTS ON TEMPORAL REALIZATION OF ESTONIAN QUANTITY

Pärtel Lippus[1,2], Juraj Šimko[2]

[1]University of Tartu, Estonia, [2]University of Helsinki, Finland
partel.lippus@ut.ee, juraj.simko@helsinki.fi

ABSTRACT

The Estonian three-way quantity system is usually described by a robust foot-level compensatory mechanism of the stressed and the unstressed syllable durations. Beneath the three-way foot-level contrast, the quantity oppositions can be realized by seven different segmental combinations of the initial vowel and the following intervocalic consonant length. In this work we investigated the effects of context dependency on durational patterns marking this 7-way contrast. We observed a strong influence of the word initial consonant and vocalic quality context on the durations of individual segment in different quantity combinations. The results show a complex relation between the intrinsic phonemic properties and suprasegmental features. Finally, we argue that the effects of segmental context and natural variation of phonetic realization might in some cases combine to be greater than average differences between quantity degrees.

Keywords: intrinsic duration, segmental context, three-way quantity, Estonian

1. INTRODUCTION

The three-way quantity system is a well-known feature of Estonian. In a number of studies (e.g. [3, 7, 8]) it has been shown that the domain of the quantity is a left-headed (i.e. stress on the first syllable) disyllabic foot. Incorporating all segmental durations within the foot, the quantity opposition between short (Q1), long (Q2) and overlong (Q3) feet is realized by the stressed-to-unstressed syllable rhyme duration ratio and some additional features (namely the pitch movement). As the unstressed syllables cannot have distinctive length oppositions independent of the stressed syllables, their duration lengthens or shortens compensatorily in the opposite direction of the stressed syllable.

While on the foot level there is a three-way opposition of stressed-to-unstressed syllable length oppositions, the stressed syllable rhyme duration is achieved by combining the length of the vowel and the coda consonant. The three-way contrast can be accomplished by lengthening the stressed vowel, the intervocalic consonant, or both, enabling minimal septets of CVCV-sequences based on seven segmental combinations: Q1: short V1 short C2 [sɑte] — Q2: long V1 short C2 [sɑːte], short V1 long C2 [sɑtte], long V1 long C2 [sɑːtte] — Q3: overlong V1 short C2 [sɑːːte], short V1 overlong C2 [sɑtːte], overlong V1 overlong C2 [sɑːtːte].

There has been some debate over the years concerning the domain of the quantity and the number of quantity degrees in Estonian, depending on the prosodic level where the feature is analyzed. On the segmental level, the difference between short, long and overlong segments is not symmetrical. Long segments are usually clearly twice the length of the short ones, but the difference between long and overlong segments is minute (albeit significant) [8], and the distinction between Q2 and Q3 is perceived by also assessing the following unstressed syllable duration [1, 3].

The syllable rhyme duration ratio is a robust way of describing the distinction, but this has been usually analyzed within similar segmental context (e.g. by limiting the study to a single minimal triplet, most frequently [sɑtɑ] – [sɑːtɑ] – [sɑːːtɑ]) [5, 9] or ignoring the segmental context by randomly pooling different phoneme combinations together [8].

On the other side, the intrinsic properties of sound segments, claimed to have rather a psychophysiological than language-specific nature [6], have shown to have non-linear effect on the quantity-related duration variability in Estonian. Eek and Meister [2, 4] have demonstrated that while [t] is shorter than [p], the ratio between the segments in Q1, Q2, and Q3 is also different: for [p] it is 1 : 2.0 : 4.5, and for [t] 1 : 2.8 : 7.0 (while for example for [ʃ] the ratio is 1 : 1.5 : 2.2) [4]. As for perception, consonant context does not have significant effect on the following vowel, but vowel quality can shift the perceived short-long category boundary earlier for [i] than for [ɑ] [10]. For closely related Finnish with

its two-way quantity contrast, rich interactions between quantity and segmental context have recently been reported in [11] and attributed to articulatory synergies and production–perception tradeoffs.

The aim of this study is to observe the interaction of the intrinsic properties of segments with the quantity in Estonian. Focussing on the context sensitivity of the effect, we show that the segmental temporal patterns are not only determined by the overall quantity level of the disyllabic foot, but they also depend of segmental attributes such as articulatory properties.

2. MATERIAL AND METHODS

The test subjects were six native Estonian speakers (3 female and 3 male, mean age 38, ranging from 24 to 57). The speakers had no reported speech or language processing disorders.

For this study we recorded the audio signal, and the tongue, jaw, and lip movements using electromagnetic articulatography (EMA) from the subjects when producing all possible quantity combinations of intervocalic bilabial stops (only acoustic data from this material are used in this paper). Also the context was altered, resulting in 4 x 7 = 28 test words listed in Tab. 1. The stimuli thus differed in quantity patterns, word initial consonant context (p vs. t), and vocalic context (a-i vs i-a).

Table 1: Stimuli for each of the seven possible quantity combinations.

Q1	[papi]	[pipa]	[tapi]	[tipa]*
Q2	[pa:pi] [pap:i]* [pa:p:i]	[pi:pa] [pip:a] [pi:p:a]	[ta:pi] [tap:i]* [ta:p:i]	[ti:pa] [tip:a]* [ti:p:a]
Q3	[pa::pi] [pap::i]* [pa:p::i]	[pi::pa] [pip::a] [pi:p::a]	[ta::pi] [tap::i]* [ta:p::i]	[ti::pa]* [tip::a] [ti:p::a]

As only 7 of the 28 test word are meaningful words in Estonian (marked with * in Tab. 1), each test word was presented to the test subjects along with a sentence where a segmentally similar word with same quantity degree was used. The test words were presented to the test subjects on a computer screen and the subjects were instructed to repeat the word about ten times without pauses between repetitions when the screen went blank. The 28 test words were randomized within a trial. To eliminate the boundary lengthening effects, the first and the last repetitions were excluded from the analysis. Three to five trials were recorded for each speaker, resulting in between 195 and 246 tokens for each word-quantity combination.

3. RESULTS AND DISCUSSION

3.1. Context effects on segmental durations

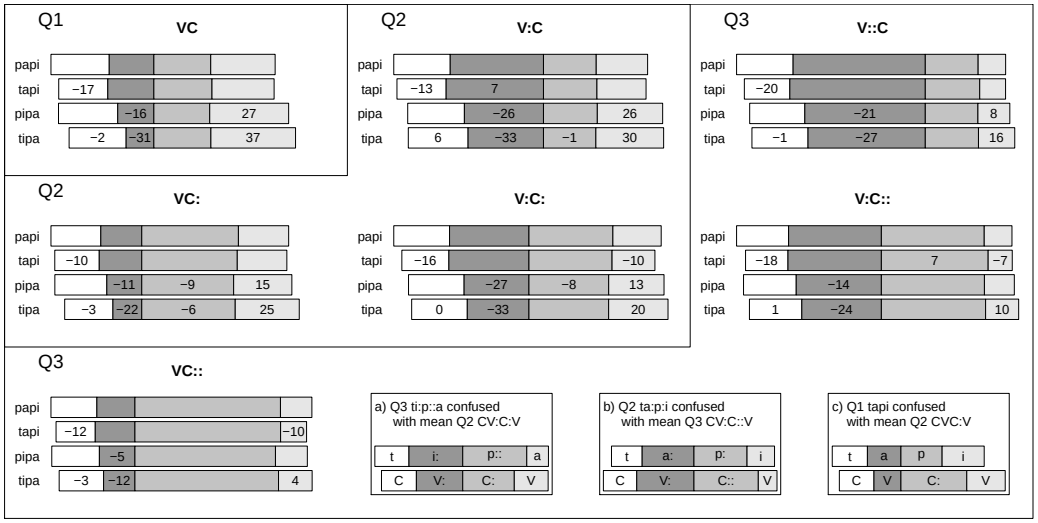
For each of the four segments in the CVCV target word, linear mixed effect models with the segment duration as a dependent variable, both vocalic context (a-i vs. i-a), and word initial consonant (p vs. t) as independent factors were fitted separately for each quantity combination; speaker influence on slopes for the independent variables was treated as a random factor. Overall, 28 models were produced.

The results are summarized in Fig. 1. For each quantity combination, *papi* is taken as the reference value (shown in the first row in each panel). The second row (*tapi*) plots the main effect of the initial consonant context, the following row, *pipa*, the main effect of vocalic context, and the last, *tipa*, the interaction of both. The numbers on the segments are the estimated differences (in milliseconds) compared to the base values and are shown only if the effect is significant.

The initial consonant context has a significant main effect on its own duration in all quantity degrees: [p] is 10–20 ms longer than [t]. However, in many cases the consonant context has a significant effect also on the second vowel duration: in *tapi*-context the final [i] is shortened roughly by 10 ms in some Q2 and Q3 quantity combinations, but the effect is not constantly significant in all quantity combinations. As the data was collected by repetitions of the token words, the word final vowel was always followed by the initial consonant of the next repetitions, thus it is likely that the consonant context affects the duration of the vowel preceded by the [t] rather than the one occurring later in the word. The consonant context had a significant effect on the following V1 only in the Q2 [ta:pi] where the long [a] was lengthened by 7 ms, and on the intervocalic consonant in Q3 [ta:p::i], where the overlong [p] was also lengthened by 7 ms.

The vocalic consonant context shows a general pattern that the [a] is longer than [i], but as both V1 and V2 play a part in the quantity-related variation, this has a complex non-linear pattern. In the case of V1, [i] is shortened more if the V1 is long, and less, if short V1 is followed by a long C2. In the case of V2, [a] is the most lengthened in Q1, where V2 is lengthened also to compensate for the V1 being short, but the intrinsic lengthening effect is the smallest or not significant at all in the case of Q3, where the V2 is quantity-wise shortened to extreme

Figure 1: Segmental durations in the seven possible quantity combinations estimated by the mixed effect models. The numbers show significant differences (in ms) compared to the reference *papi* ($p < 0.05$). The segments are aligned by the onset of the medial consonant. The insets (a–c) (bottom right) give three examples of possible durational overlap.



to compensate for the stressed syllable being extra long. The vocalic consonant context has no effect on the word initial consonant duration.

To some extent, however, the interaction of the consonant and the vocalic contexts seems to cancel out the main effect of the consonant context. Comparing *tipa* to *papi*, the difference between initial [t] and [p] duration is only about 3 ms. The vocalic context effect, on the other hand, is even stronger in case of the interaction: the [i] in V1 position is shortened by an extra 5–15 ms in *tipa*-context compared to *pipa*-context, while [ɑ] in V2 position is lengthened by 4–10 ms. Again, the contextual effects on the C2 duration are the smallest, being significant only in Q2, where in [tipːɑ] the [p] is shortened by 6 ms.

Different intrinsic duration of segments and their different sensitivity to lengthening effects might lead to considerable overlaps in durational patterns pertaining to distinct quantity combinations. The insets in Fig. 1 illustrate this point. Panel Fig. 1a) compares the mean durational pattern of a Q3 [tiːpːːɑ] with a mean pattern of Q2 CVːCːV quantity combination. The naturally shorter duration of [iː] compared to [ɑː] combined with the additional shortening due to [t] context reduced the expected Q2–Q3 difference in V1 duration. At the same time, the lengthening of word final [ɑ] due to interaction with initial [t] further reduces the overall difference between duration patterns. As a result, at least when it comes to segment durations, Q3 [tiːpːːɑ] is not considerably different from an average token from Q2 CVːCːV category. Similar trends and consequences can be observed in the other two cases comparing Q2 and Q1 words with mean patterns from Q3 and Q2 categories, respectively.

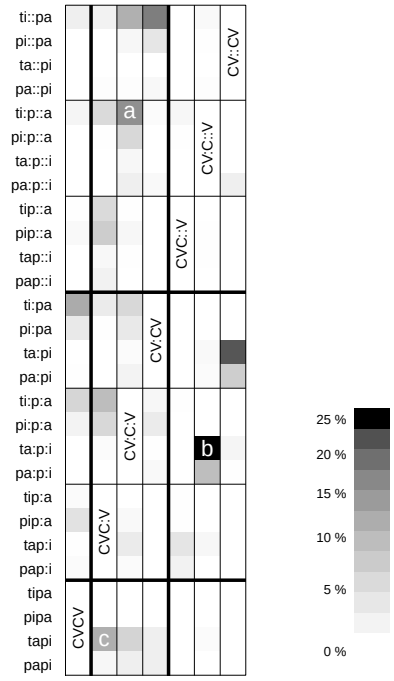
3.2. Overlaps of overall duration patterns

In the light of these phenomena and given the natural variance ubiquitous in speech production, there arises a question whether the durational effects of quantity are robust enough to maintain required contrast in spite of these complex interactions arising from segmental context. In order to address this question, we estimate the likelihood of possible overlaps between durational patterns of word tokens differing in both segmental content and quantity, as suggested in Fig. 1.

To capture the duration of all segments in our words in parallel, we can conceptualize each token as a point in a 4-dimensional space with axes representing segment durations at the four subsequent positions: C1, V1, C2 and V2. In this space, points representing individual tokens from the same quantity category should form a bounded region. The assumption of segmental duration alone providing a sufficient contrast is equivalent to an expectation of no or minimal overlaps among these seven regions.

For each speaker, we first found the median position of token-points belonging to each quantity-

Figure 2: Frequency of tokens with duration patterns nearer to patterns of quantity combination other than their own.



combination region (its coordinates are median durations at four positions). Subsequently, we calculated the Euclidean distance of every point from each of these 7 medians. Finally, for each point we evaluated the percentage of how many times it ended up too close to a "wrong" category center: we counted all occasions when a token-point was actually closer to the median of a different quantity-combination category than to its own one.

Fig. 2 displays the result of this procedure in a form of a customized confusion matrix. The results of all 6 speakers are summed in the picture. The shade of squares shows the frequency of occurrence of the event when a the durational pattern of a token (indicated by the row name) is close to the in-correct quantity combination median (labeled at the diagonal). The darker the shade, the higher the frequency of such incident. Also in the figure, the examples discussed above and shown in Fig. 1 are marked.

As seen in Fig. 2, at least in the light of this type of production analysis, a small amount of durational overlaps is present for most combinations of segmental context and quantity. It also shows a rather robust pattern associated with vocalic context: overall duration patterns of *a–i* tokens is closer to that of median patterns of the higher quantity while *i–a* tokens are more similar to the lower quantity medians. This is in line with our analysis of intrinsic durations and the effects of quantity on vowel durations.

Within the Q3 region (top right corner), the durational overlaps are relatively rare compared to Q2. Due to the interactions between initial consonant and vocalic context, the overlaps are considerably more frequent for tokens with initial [t].

The analysis presented here offers several insights into parsing quantity contrast in Estonian. It is important to note that we are not suggesting that the effects of context dependency necessarily have detrimental consequences on listeners' ability to correctly evaluate contrast between quantity combinations with overlapping duration patterns. What our results do imply, however, is that the listeners most likely process the primary temporal cues of quantity contrast in a context-dependent manner. Moreover, the segmental context that is presumably taken into account is not limited to the locus of quantity marking (the segments within the disyllabic foot in our case) and includes further contextual factors.

Two possible extensions of this work offer themselves naturally. A perception study focusing on "difficult" combinations identified here can shed further light on quantity processing. On the other hand, an articulatory analysis of the material (currently on its way) can reveal the underlying articulatory sources of the observed interactions.

4. CONCLUSIONS

In this work we investigated the effect of segmental context on the complex quantity system of Estonian. We have identified several influences of segmental quality on durational patterns of seven quantity combinations present in the system. The interactions are not linear nor are they limited to the effects of intrinsic phoneme durations; they also involve the place of articulation of the word initial consonant. The durational differences marking quantity contrast are in some cases not robust enough to make the contextual effects negligible. Therefore, our results strongly suggest that further analysis taking segmental context into account could provide new impetus for research of phonological quantity.

5. ACKNOWLEDGEMENTS

We express our gratitude to Mona Lehtinen for the help with recording the data, Helen Türk for acoustic labeling of the material, and to all participants. The work was partly funded by the Estonian Research Council grant IUT2-37 for the first author and the Finnish Academy grant to the second author.

6. REFERENCES

[1] Danforth, D. G., Lehiste, I. 1977. Foneettisten vihjeiden hierarkia viron kvantiteetin havaitsemisessa. *Virittäjä* 81(4), 404–411.

[2] Eek, A. 1974. Observations on the duration of some word structures: I. In: Eek, A., (ed), *EPP 1974. To Honour Academian Paul Ariste on the Occasion of his Seventieth Birthday, February 3, 1975* Estonian Papers in Phonetics. Tallinn: Academy of Sciences of the Estonian S.S.R. 18–32.

[3] Eek, A., Meister, E. 1997. Simple perception experiments on estonian word prosody: Foot structure vs. segmental quantity. Lehiste, I., Ross, J., (eds), *Estonian prosody: Papers from a symposium* Tallin, Estonia. Institute of Estonian Language 71–99.

[4] Eek, A., Meister, E. 1999. Estonian speech in the babel multi-language database: Phonetic-phonological problems revealed in the text corpus. Fujimura, O., Palek, B., (eds), *Proceedings of LP* volume 98 Prague. The Karolinum Press 529–546.

[5] Eek, A., Meister, E. 2003. Domain of the estonian quantity degrees. evidence from words containing diphthongs. Solé, M.-J., Recasens, D., Romero, J., (eds), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* 2039–2042.

[6] Lehiste, I. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.

[7] Lehiste, I. 2003. Prosodic change in progress: From quantity language to accent language. In: Fikkert, P., Jakobs, H., (eds), *Development in prosodic systems*. Berlin/New York: Mouton de Gruyter 47–65.

[8] Lippus, P., Asu, E. L., Teras, P., Tuisk, T. 2013. Quantity-related variation of duration, pitch and vowel quality in spontaneous estonian. *Journal of Phonetics* 41(1), 17–28.

[9] Lippus, P., Pajusalu, K., Allik, J. 2009. The tonal component of estonian quantity in native and non-native perception. *Journal of Phonetics* 37(4), 388–396.

[10] Meister, E., Werner, S., Meister, L. 2011. Short vs. long category perception affected by vowel quality. In: Lee, W.-S., Zee, E., (eds), *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong 1362–1365.

[11] Šimko, J., O'Dell, M., Vainio, M. 2014. Emergent consonantal quantity contrast and context-dependence of gestural phasing. *Journal of Phonetics* 44, 130–151.