Variation in Vowel Quality as a Feature of Estonian Quantity

Pärtel Lippus

Institute of Estonian and General Linguistics, University of Tartu, Estonia
partel.lippus@ut.ee

Abstract

The three-way distinction of Estonian quantity is a feature of a primary stressed disyllabic foot. Quantity degrees are realized by the combination of segment duration ratios and the pitch contour within the foot. Additionally, other phonetic features appear to follow a similar pattern, such as vowel quality. In this study, segment duration and vowel quality were analyzed in Estonian spontaneous speech. Vowels in stressed syllables of short quantity degree feet were distinctively closer to the center compared to those of long and overlong quantity degree feet. Vowels in unstressed syllables showed more variation in general, but the difference between the quantity degrees was relatively small.

Index Terms: Estonian, quantity, vowel quality

1. Introduction

In Estonian, disyllabic feet can be in short (Q1), long (Q2), or overlong (Q3) quantity degree. The feet are left headed and phonologically it is the stressed vowel (e.g. [vilu˙] 'chilly' – [vi:lu] 'slice, sg. gen.' – [vi::lu] 'slice, sg. part.'), consonants between syllables (e.g. [kɑlʲi˙] 'kvass' – [kɑlʲli] 'hug, sg. nom.' – [kɑlʲ:li] 'precious, sg. gen.'), or a combination of a stressed vowel and following consonant (e.g. [sɑte˙] 'fall-out' – [sɑ:tte] 'get, pl. 2nd pers.' – [sɑ:t:te] 'broadcast, sg. gen.') that carries the quantity, whereas the following unstressed syllables do not have length opposition [20]. As there is a tendency to foot isochrony [16], [17], the duration of the second syllable compensates for the variation of the first syllable. The duration of syllable onset consonants are mainly dependent of the local speaking rate while the quantity can be described as the ratio of syllable rhyme duration, e.g. [1], [7], [8], [10], [11], [12], [17], or by comparing the V1 duration with the weighted sum of segment durations within a foot, e.g. [9], [19].

While the temporal structure of a disyllabic foot is the primary feature of quantity, variation of the pitch contour has also been studied extensively. Typically the pitch is relatively flat in the first syllable and falls at the syllable boundary in Q1 and Q2, whereas in Q3 pitch falls at the beginning of the first syllable [1], [10], [11], [12]. It is thought that the pitch cue is of vital importance for discriminating Q2 and Q3 [5], [11]. On the other hand this claim has been doubted as there are many cases where most of the pitch contour in the first syllable is interrupted by a voiceless consonant (e.g. [kɑt:tɑ]) [19]. Nevertheless, recent perception studies [13], [14] show that conflicting temporal and pitch cues can confound the discrimination of Q2 and Q3, whereas temporal cues are sufficient for successful discrimination if the pitch cue is not present. The weight of the pitch cue appears to vary in accordance with the dialectal background of the listeners [13]. These results suggest that instead of a fixed set of features that describe the quantity degrees, there is a more complex interaction between different features that are weighed by the listener.

In Estonian there are nine vowels /i, y, u, e, ø, ɤ, o, æ, ɑ/ that can occur in the first syllable, but only four of them /i, u, e, ɑ/ can occur also in non-initial syllables, e.g. [2]; sometimes the low front vowel is marked with /a/ instead of /æ/, e.g. [20]. Vowel quality has been shown to vary in connection with quantity; though vowel length has a relatively small effect on vowel quality. In the stressed syllable, vowels in Q3 feet are the most peripheral (while longest in duration) while vowels in Q1 feet are the most centralized. This variation, however, does not exceed 1 Bark difference and therefore is not considered to be perceivable [6]. Vowel quality perception in Estonian has been assumed to be unrelated to vowel duration, because as in other quantity languages it is used for quantity opposition. However, a recent study shows that also changes in segment duration do affect the perception of vowel quality [15]. In an unstressed syllable quantity degrees affect the vowel quality in the opposite direction and the variation crosses the perceptual boundaries: vowels of Q1 feet are the most peripheral (and the longest in duration) and the vowels of Q3 feet are the most centralized [6].

2. Materials and methods

The data were extracted from the University of Tartu phonetic corpus of Estonian spontaneous speech. The corpus consists of 29 hours of spontaneous dialogues and monologues from 35 speakers. The speech is manually segmented at the word and segmental level (work in progress; currently about 21 hours completed). For this paper 11 hours and 44 minutes of speech from 14 speakers was used.

All the 14 subjects (6 female, 8 male; age ranging from 21 to 50 years with an average of 33.8 years) are native Estonian speakers with university education who live in Tartu or Tallinn. The original regional background of the speakers shows more variation: only two of the speakers are grown up in Tallinn and four in Tartu, four in different villages in central Estonia, one in West Estonia, and three in South Estonia.

Table 1. *Number of observations of vowels in the first and the second syllable.*

		V1						V2			
Sex	**Foot**	**i**	**u**	**e**	**o**	**æ**	**ɑ**	**i**	**u**	**e**	**ɑ**
F	Q1	32	16	33	17	27	29	10	45	16	83
	Q2	11	5	32	9	8	9	14	3	44	13
	Q3	7	11	14	9	5	12	17	1	12	28
M	Q1	56	23	43	24	44	48	35	55	27	121
	Q2	27	10	29	25	14	21	25	2	82	17
	Q3	11	9	14	13	6	23	24	0	12	40

The data were analyzed with Praat [4]. Words with two open syllables in Q1, Q2 and Q3 were found and segment durations as well as F1, F2, F3, and F4 values at the mid-point of V1 and V2 were extracted with a Praat script. Formant values found by the script were manually checked.

The vowels /y, ø, ʏ/ and the unstressed vowel /u/ were left out of the analysis because there were less than five observations of each vowel in the first syllable of Q2 and Q3 words. In total 726 words were analyzed. The number of tokens for each vowel is presented in Table 1. There were 392 words in Q1, 200 words in Q2, and 134 words in Q3.

The segment durations and the vowel formant values were compared between the quantity degrees. A one-way ANOVA was used to test the difference of each characteristic. In order to plot the vowels in two formant space, F2´ was calculated using the formula from [3]. The formant values were converted to Bark using the formula from [18]. A multinomial logistic regression model was used to evaluate the influence of the acoustic characteristics to the quantity degrees.

3. Results and Discussion

The mean segment durations are presented in Figure 1. The duration of syllable onset consonants varies slightly between the quantity degrees. The duration of initial consonants (C1) is 57 ms in Q1, 77 ms in Q2, and 79 ms in Q3. The difference between Q1 vs. Q2 and Q3 is about 20 ms (by ANOVA the difference is significant $F(2) = 57.79$, $p < 0.001$). The duration of intervocalic consonants (C2) is 58 ms in Q1, 56 ms in Q2, and 67 ms in Q3, i.e. the difference between Q1 and Q2 vs. Q3 is about 10 ms ($F(2) = 12.754$, $p < 0.001$). This variation can be a result of the local speech rate, and could be reduced by taking into account the position of the word in the phrase, as was done in [1], a study using the data from the same corpus. The reason why this was not done here is that there were too few observations of each vowel in each possible phrasal position.

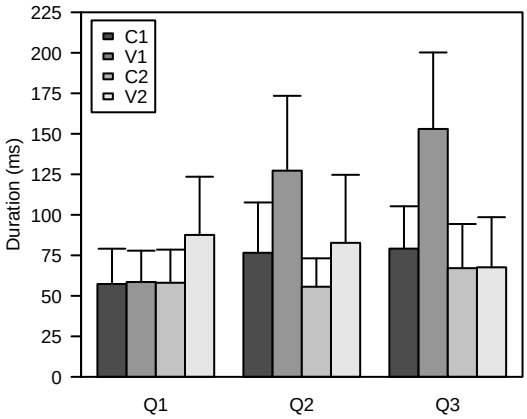

Figure 1: *Mean segment duration and standard deviation.*

The mean duration of the stressed vowel (V1) is 58 ms in Q1, 127 ms in Q2, and 153 ms in Q3, e.g. V1 is about twice as long in Q2 than it is in Q1, but only about 20 % longer in Q3 than in Q2 ($F(2) = 492.42$, $p < 0.001$). The mean duration of unstressed vowels (V2) is 88 ms in Q1, 83 ms in Q2, and 68 ms in Q3 ($F(2) = 14.695$, $p < 0.001$).

The ratio of syllable rhymes was calculated (see Figure 2). The S1/S2 ratio is 0.7 in Q1, 1.7 in Q2, and 2.5 in Q3 ($F(2) = 455.29$, $p < 0.001$). Despite the relatively high deviations these results are similar to those found in earlier studies, e.g. [1], [11].

In Figure 3 the vowels are plotted in the space of F1 and F2´. Stressed vowels are closer to the center in Q1 and more peripheral in Q2 and Q3, but there is not much difference in V1 quality between Q2 and Q3. Especially for female speakers, high vowels are more centralized in front-back direction and low vowels are more centralized in high-low direction. The difference between Q1 vs. Q2 and Q3 exceeds 1 Bark level.

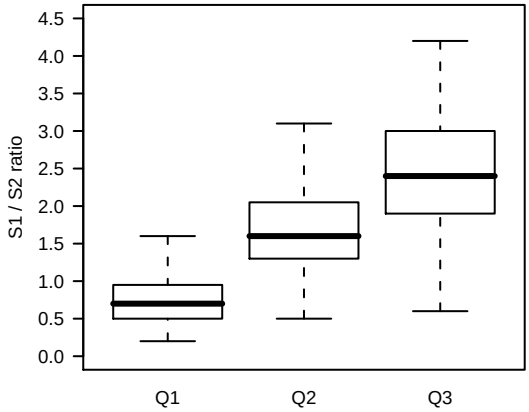

Figure 2: *The duration ratio of syllable rhymes.*

Unstressed vowels show more variation in general, but this variation is less related with the quantity. In [6] it has been noted that there is a tendency of unstressed /e/ of a Q1 foot changing into /æ/. It seems that the place of articulation of an unstressed /e/ is that of the low front vowel /æ/ in feet of all quantity degrees. In Q3 feet, the /e/ of female speakers is somewhat higher than in Q1 and Q2 feet. The unstressed vowel in Q3 is usually shortest in duration, so the higher F1 can be considered as movement towards the schwa /ə/. Differences of the mean values of unstressed /i/ and /ɑ/ formants do not exceed 1 Bark between quantity degrees. Unfortunately there were too few observations of the vowel /u/ in the unstressed syllable of Q2 and Q3 feet to analyze them.

In order to compare vowel quality in relation to the quantity of the word, formant values were normalized by calculating the ratio of single vowel formant values and the mean values of that vowel. As vowel reduction is directed toward the center of the vowel space, ratios were calculated for the following relationships: $F1_i/F1_{mean}$ for low vowels, $F1_{mean}/F1_i$ for the high vowels, $F2_i/F2_{mean}$ for front vowels, $F2_{mean}/F2_i$ ratio for back vowels. Therefore, the ratio is more than 1 for more peripheral formant values and less than 1 for more centralized formant values.

Formant ratios of V1 both for F1 and F2 are 0.9 in Q1 foot and 1.0 in Q2 and Q3 feet. The difference of Q1 vs. Q2 and Q3 is significant for F1 ratio at $F(2) = 24.440$, $p < 0.001$ and for F2 ratio at $F(2) = 122.24$, $p < 0.001$. The mean F1 ratio of V2 is 1.0 in all cases (1.020 in Q1, 1.008 in Q2, and0.983 in Q3), but possibly due to the higher F1 value of /e/ in Q3 feet of female speakers, an ANOVA finds a significant difference between quantity degrees ($F(2) = 3.326$, $p < 0.01$). The mean F2 ratio of V2 is also 1.0 in all cases and there is no variation between the quantity degrees ($F(2) = 1.453$, $p = 0.235$).

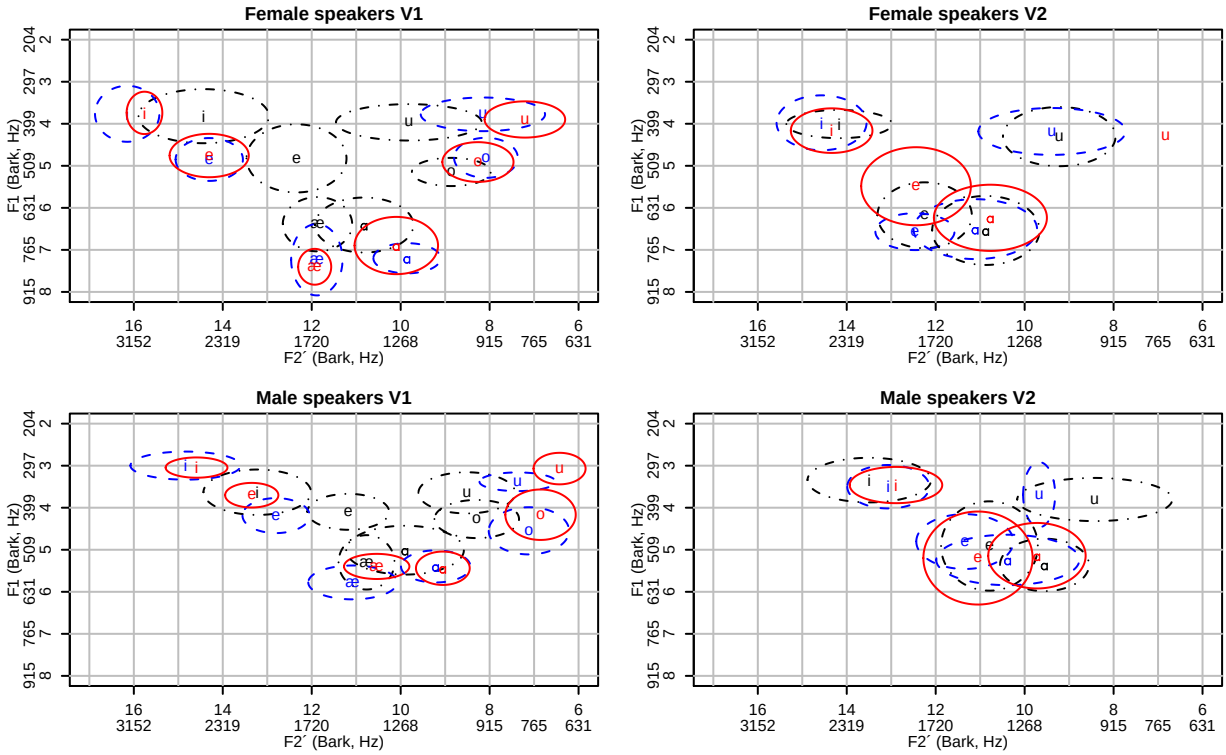

Figure 3: *Stressed and unstressed vowels of male and female speakers in the space of F1 and F2´. The standard deviation is plotted with an ellipse as follows: vowels from Q1 feet are plotted in dot-dashed black lines, vowels from Q2 feet in slashed blue lines, and vowels from Q3 feet in solid red lines.*

Table 2. *Multinomial logistic regression analysis of the quantity of the foot (Q3 as the reference level).*

	b	S.E.	z-value	Pr(>|z|)	exp(b)
Q1 vs. Q3					
Intercept	20.214	3.786	5.339	<0.001	
C1 duration	-0.027	0.009	-3.053	<0.005	0.973
V1 duration	-0.156	0.015	-10.161	<0.001	0.856
C2 duration	0.042	0.014	2.986	<0.005	1.043
V2 duration	0.076	0.009	8.627	<0.001	1.079
V1 F1 ratio	-7.306	2.168	-3.370	<0.001	0.001
V1 F2 ratio	-7.370	2.709	-2.720	<0.01	0.001
V2 F1 ratio	4.628	1.760	2.629	<0.01	102.324
V2 F2 ratio	-1.427	2.310	-0.618	0.268	0.240
Q2 vs. Q3					
Intercept	5.388	2.310	2.333	<0.01	
C1 duration	-0.004	0.005	-0.974	0.165	0.996
V1 duration	-0.023	0.004	-5.699	<0.001	0.977
C2 duration	-0.030	0.007	-4.357	<0.001	0.971
V2 duration	0.035	0.005	6.447	<0.001	1.035
V1 F1 ratio	-2.474	1.227	-2.017	<0.05	0.084
V1 F2 ratio	-0.351	1.531	-0.229	0.409	0.704
V2 F1 ratio	1.329	1.054	1.261	0.104	3.776
V2 F2 ratio	-0.666	1.330	-0.501	0.308	0.514

Finally, a multinomial logistic regression model was created to describe the effect of the segment duration and the formant ratios on the quantity of the foot. The output of the model is presented in Table 2.

In the comparison of Q1 vs. Q3 all segment durations are significant, but the most significant results are the effects of V1 and V2 duration. The formant ratios of V1 are also significant, but we have to keep in mind that these values are ratios, so a 0.1 unit change would give an -10% impact. Surprisingly, the F1 formant ratio of V2 has a significant impact, even thought the differences of the mean values are minimal. The impact of V2 F2 ratio is not significant.

In the comparison of Q2 vs. Q3, the duration of C1 is not significant and the effects of V1 and V2 are much weaker. Also, the F1 ratio of V1 is significant, even though the mean difference of the ratio between Q2 and Q3 is relatively small. The F2 ratio of V1 and the formant ratios of V2 are not significant.

Table 3. *The probability of the quantity predicted by the model using the mean values.*

	p_{Q1}	p_{Q2}	p_{Q3}
Q1 mean values	0.992	0.008	0.000
Q2 mean values	0.002	0.760	0.238
Q3 mean values	0.000	0.401	0.599

In order to evaluate the goodness of this model, the mean values of independent variables were used to predict the quantity (see Table 3). The model seems to handle the

opposition of Q1 vs. Q2 and Q3 very well, but the opposition of Q2 and Q3 is not so clear: with the mean values of Q2 the model predicts 76% Q2 vs. 24% Q3 and with the mean values of Q3 it predicts 60% Q3 vs. 40% Q2.

A lot of variability in segment duration could be reduced by taking the phrasal position and accentuation conditions into account. As was previously mentioned, this was not done in this study because it would have produced combinations of variables to which there were no corresponding observations. The model could be improved also by considering a characteristic of the pitch contour as a variable.

4. Conclusions

As has been found in previous studies, the duration of vowels is more important for Estonian quantity opposition than the duration of syllable initial consonants. Rather than the V1 duration by itself, it is the ratio of the segment durations within the foot that describes the quantity degrees contrastively.

The variation in vowel quality is related to the quantity. Vowels in stressed syllables of Q1 feet are closer to the center and in stressed syllables of Q2 and Q3 feet they are more peripheral. The difference in V1 quality between Q1 vs. Q2 and Q3 should be perceivable as it exceeds 1 Bark difference. Vowels in unstressed syllables vary significantly, but the most of the variation is not connected with the quantity of the foot. While the space of V2 in general is more centralized, the vowel /e/ has moved to the low front corner of the space, and is realized as /æ/.

5. Acknowledgements

The present research was supported in part by the Estonian Science Foundation Grant no. 7904. The author would like to thank Eva Liina Asu, Ilse Lehiste, and Karl Pajusalu for their useful comments, Ene Käärik and Stefan Werner for their advice on statistics, and Cameron Robert Rule for editing the language of this paper.

6. References

[1] Asu, E. L., Lippus, P., Teras, P. & Tuisk, T. (2009). The realization of Estonian quantity characteristics in spontaneous speech. In Vainio, M., Aulanko, R., Aaltonen, O. (Eds.) *Nordic Prosody. Proceedings of the Xth Conference, Helsinki 2008.* (pp. 49-56), Peter Lang, Frankfurt.

[2] Asu, E. L. & Teras, P. (2009). Illustrations of the IPA. Estonian. *Journal of the International Phonetic Association, 39/3,* 367-372.

[3] Bladon, A. & Fant, G. (1978). A two-formant model and the cardinal vowels. *STL-QPSR, 1,* Stockholm, 1-8.

[4] Boersma, P. & Weenink, D. (2009). *Praat: doing phonetics by computer* (Version 5.1.15) [Computer program]. Retrieved August 30, 2009, from http://www.praat.org/

[5] Eek, A. (1980). Further information on the perception of Estonian quantity. *Estonian Papers in Phonetics 1979,* 31-57.

[6] Eek, A. & Meister, E. (1998). Quality of standard Estonian vowels in stressed and unstressed syllables of the feet in three distinctive quantity degrees. *Linguistica Uralica, 3 (XXXIV),* 226-233.

[7] Eek, A. & Meister, E. (2003). Foneetilisi katseid ja arutlusi kvantiteedi alalt (I). Häälikukestusi muutvad kontekstid ja välde. *Keel ja Kirjandus, 11-12,* 815-837, 904-918.

[8] Eek, A. & Meister, E. (2004). Foneetilisi katseid ja arutlusi kvantiteedi alalt (II). Takt, silp ja välde. *Keel ja Kirjandus, 4-5,* 251-271, 336-357.

[9] Krull, D., Traunmüller, H. & Bertinetto, P. M. (2006). Local speaking rate and perceived quantity: an experiment with Italian listeners. *Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics Working Papers 52,* 81-84.

[10] Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. *American Studies in Uralic Linguistics, 1* (pp. 21-82), Bloomington.

[11] Lehiste, I. (2003). Prosodic change in progress: from quantity language to accent language. In Fikkert, P., Jacobs, H. (Eds.) *Development in Prosodic Systems.* (pp. 47-66), Studies in Generative Grammar 58. Mouton de Gruyter, Berlin, New York.

[12] Liiv, G. (1961). Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid. *Keel ja Kirjandus, 7-8,* 412-424, 480-490.

[13] Lippus, P. & Pajusalu, K. (2009). Regional variation in the perception of Estonian quantity. In Vainio, M., Aulanko, R., Aaltonen, O. (Eds.) *Nordic Prosody. Proceedings of the Xth Conference, Helsinki 2008.* (pp. 151-157), Peter Lang, Frankfurt.

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

[15] Meister, E. & Werner, S. (2009). Duration affects vowel perception in Estonian and Finnish. *Linguistica Uralica, 3 (XLV),* 161-177.

[16] Nolan, F. & Asu, E. L. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica, 66,* 64-77.

[17] Ross, J. & Lehiste, I. (2001). *The temporal structure of Estonian runic songs*. Mouton de Gruyter, Berlin.

[18] Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *J. Accoust. Soc. Am. 88,* 97—100.

[19] Traunmüller, H. & Krull, D. (2003). The effect of local speaking rate on the perception of quantity in Estonian. *Phonetica, 60,* 187-207.

[20] Viitso, T.-R. (2003). Phonology, morphology and word formation. Erelt, M. (Ed.) *Estonian Language.* (pp. 9-92), Linguistica Uralica, Supplementary Series 1. Estonian Academy Publishers, Tallinn.