Journal of Phonetics 41 (2013) 17–28

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian

Pärtel Lippus [a,*], Eva Liina Asu [a], Pire Teras [a,b], Tuuli Tuisk [a]

[a] Institute of Estonian and General Linguistics, University of Tartu, Ülikooli 18, Tartu 50090, Estonia
[b] Department of Finnish, Finno-Ugrian and Scandinavian Studies, University of Helsinki, Finland

ARTICLE INFO

Article history:
Received 17 April 2011
Received in revised form
11 September 2012
Accepted 19 September 2012
Available online 27 November 2012

ABSTRACT

The three-way distinction of the Estonian quantity is a feature of the primary stressed disyllabic foot. The quantity degrees are realized by different temporal patterns of the segments within the foot. Additionally, other phonetic features appear to vary depending on the quantity, such as pitch contour and vowel quality. In this paper, the quantity-related variation of segment duration, pitch, and vowel quality was investigated in spontaneous Estonian in words occurring in sentence-medial position. While a disyllabic foot can be made up of a number of segmental combinations, the data showed that the quantity opposition operates independently of the segmental structure of the foot. The effect of the analyzed features was evaluated with a multinomial logistic regression model. The model showed that all the variables which were included had a significant effect. Besides the inverse relation of the stressed and unstressed syllable rhymes, which is traditionally used to describe the three-way opposition, syllable onset duration also had a weak effect. Additionally, vowel quality was significant for the opposition of short (Q1) and long (Q2), and pitch alignment was significant for the opposition of long (Q2) and overlong (Q3) quantities.

1. Introduction

Estonian is one of the few languages that differentiate between three degrees of length, usually referred to as short (Quantity 1), long (Quantity 2) and overlong (Quantity 3). Three degrees of length have also been reported in other Finno-Ugric languages such as Livonian (Lehiste et al., 2008), Saami (Bye, 1997; Bye, Sagulin, & Toivonen, 2009; McRobbie-Utasi, 2007), and Soikkola Ingrian (Markus, 2011). Unrelated languages with three-way oppositions include Western Nilotic languages Dinka (Andersen, 1987; Remijsen & Gilley, 2008) and Shilluk (Remijsen, Ayoker, & Mills, 2011), dialects of Mixe in Oaxaca, Mexico (Hoogshagen, 1959; van Haitsma & van Haitsma, 1976), and the native American language Yavapai (Thomas, 1992; Thomas & Shaterian, 1990), as well as dialects of German (Bremer, 1927; Prehn, 2012) and Applecross Gaelic (Ternes, 1973).[1]

In all the above mentioned non-Finno-Ugric languages, the three-way length contrast does not include consonants but operates only on vowels in monosyllabic words. In Estonian, however, the segmental quantity opposition in monosyllabic words is binary, involving a short vowel followed by a long consonant (or consonant cluster) or a long vowel (or diphthong) followed by a short consonant. Words consisting of a single open syllable do not participate in the quantity opposition. It is only in words longer than one syllable that we find the three-way opposition in Estonian. In disyllabic words, the opposition can be carried, for instance, by the vowel of the primary stressed syllable (e.g. Q1 [vilu] 'chilly'—Q2 [viːlu] 'slice, sg. gen.'—Q3 [viːːlu] 'slice, sg. part.'), consonants on the syllable boundary (e.g. Q1 [kɑlʲi] 'kvass'—Q2 [kɑlʲli] 'hug, sg. nom.'—Q3 [kɑlʲːli] 'precious, sg. gen.'), or a combination of a stressed vowel and the following consonant (e.g. Q1 [sɑte] 'fallout'—Q2 [sɑːtte] 'get, 2nd pers. pl.'—Q3 [sɑːtːte] 'broadcast, sg. gen.'; for more examples see Asu & Teras, 2009; Viitso, 2003).

The question of the domain of the quantity has been a much debated issue in the literature on the Estonian quantity (Eek & Meister, 2003). It may be analyzed phonologically as the duration of the stressed syllable. On the other hand, it can be argued that the three-way quantity distinction in Estonian is fundamentally different from those of unrelated languages in that it operates over a disyllabic foot. Unstressed syllables do not exhibit length opposition, but due to a certain degree of foot isochrony (e.g. Lehiste, 2003; Nolan & Asu, 2009), the duration of the second syllable compensates for the durational variation in the first syllable, which means that the unstressed second syllable is the longest in Q1 and

* Corresponding author. Tel.: +372 737 6512; fax: +372 737 5224.
E-mail addresses: partel.lippus@ut.ee (P. Lippus), eva-liina.asu@ut.ee (E.L. Asu), pire.teras@ut.ee (P. Teras), tuuli.tuisk@ut.ee (T. Tuisk).

[1] For a detailed overview of the languages with a three-way contrast on vowels see Prehn (2012).

0095-4470/$ - see front matter
http://dx.doi.org/10.1016/j.wocn.2012.09.005

the shortest in Q3. Although durational variation in unstressed syllables is not phonological, it is nevertheless important for the perception of quantity—for instance, it has been shown that the opposition of Q2 and Q3 is not perceived if only the first syllable of a disyllabic sequence is presented to the listeners (Eek & Meister, 2003). Therefore, one of the main features characterizing the Estonian quantity opposition is the duration ratio of the stressed and unstressed syllable in the foot (Eek, 1974; Lehiste, 1960, 1997, 2003; Liiv, 1961).

It is not uncommon in languages with three degrees of vowel duration to sustain the contrast with the use of additional prosodic features (McRobbie-Utasi, 2007). In the Estonian quantity system, pitch is additionally employed in order to distinguish between Q2 and Q3. The fundamental frequency (F0) contour in Estonian Q1 and Q2 is realized differently than in Q3 (when produced with an H*+L pitch accent). In Q1 and Q2, there is an F0 step-down between the end of the first syllable nucleus and the beginning of the second syllable, while in Q3, a fall takes place early during the first syllable (e.g. Asu, Lippus, Teras, & Tuisk, 2009; Lehiste, 1960; Liiv, 1961; Remmel, 1975). The account of tonal characteristics of Estonian quantities along these lines goes back to the 1960s.

The description is usually in terms of the location of an F0 peak in relation to the syllable boundary, while the syllable duration is dependent on the quantity degree. What is normally not mentioned is that the absolute duration of the peak location is very similar between the quantities, being roughly at about 100 ms from the beginning of the stressed vowel (Asu et al., 2009; Lippus & Ross, 2011). Having said that, pitch alignment has been shown to be crucial for distinguishing perceptually between Q2 and Q3 (Eek, 1980; Fox & Lehiste, 1987, 1989; Lehiste, 1975; Lehiste & Danforth, 1977; Lippus, Pajusalu, & Allik, 2009). Perception tests using stimuli with manipulated segment durations have shown that Q3 can only be perceived if the step-down in F0 is located in the middle of the stressed syllable as opposed to at the syllable boundary (Eek, 1980; Lehiste, 1975, 1997, 2003; Lippus et al., 2009). Additionally, it was shown that it is possible to change Q2 vs. Q3 perception by changing only the pitch contour of a Q2 foot without manipulating the segmental duration (Lippus, Pajusalu, & Allik, 2011). On the other hand, the pitch is still a secondary cue for distinguishing between Q2 and Q3, its perceptual weight depending, for instance, on the listener's dialectal background (Lippus & Pajusalu, 2009). Nevertheless, the connection between tonal and temporal features seems to be much more general. Longer units tend to be accompanied by more dynamic pitch movements in quantity languages, while the units with more dynamic pitch movements tend to be longer in tone languages (Järvikivi, Vainio, & Aalto, 2010; Vainio, Järvikivi, & Aalto, 2010; Yu, 2010). The link between perceived duration, tone, and vowel quality is not only claimed to exist in quantity and tone languages but is a more universal phenomenon (Lehiste, 1976; Lehnert-LeHouillier, 2010).

In Estonian, vowel quality has been shown to vary in connection with quantity, although vowel duration has a relatively small effect on the quality of the vowel. While the stressed vowels in Q3 feet and the unstressed vowels in Q1 feet are longest in duration, they are the most peripheral in quality. The stressed vowels in Q1 feet and unstressed vowels in Q3 feet are shortest and most centralized (Eek & Meister, 1998). Lippus (2010) shows, however, that in spontaneous speech the quality difference between short and long vowels is considerably larger than in lab speech analyzed by Eek and Meister (1998), exceeding the level of one Bark. Consequently, this difference is significant for evaluating the subjective duration, and therefore should be included in the description of the quantity system. A recent study shows that a link between spectral and temporal cues also exists in Estonian: the perception of the short vs. long category boundary is affected by vowel quality (Meister, Werner, & Meister, 2011). At the same time, micro-durational variation has been shown to affect vowel category perception (Meister & Werner, 2009).

Most studies on quantities are based on controlled speech. The results of some earlier work on the realization of the Estonian quantity degrees in spontaneous speech have shown that both temporal and tonal characteristics of disyllabic feet appear to be stable in spontaneous data. According to Krull (1993a, 1993b, 1997, 1998) the most stable acoustic difference between the quantities in spontaneous speech is the duration ratios, but in her data she does not observe any stability of tonal correlates. The latter is probably due to the fact that conditions such as the position of the word in the utterance, accentuation, and intonation were not factored in the analysis. However, following research on Estonian intonation (Asu & Nolan, 1999, 2007), which demonstrates that the pitch cue to the quantity is sensitive to the intonational category, Asu et al. (2009) show that if the above listed conditions are taken into account, the realization of the tonal characteristics of the three quantities is also relatively stable in spontaneous speech. Thus, in their data, the location of the pitch peak in relation to the beginning of the first vowel (V1) was shown to be stable.

Typologically the opposition of three degrees of length is a rare phenomenon in the world's languages, and it is therefore not surprising that phonological treatments favor a binary analysis to the ternary one. In theory, the Estonian quantity system yields to both types of analysis. A number of studies conclude that the Estonian quantity contrast is evolving into a system where the opposition of Q1 and Q2 includes short vs. long segments, but the opposition of Q2 and Q3 involves additional features such as pitch and syllable ratios (Eek, 1980; Lehiste, 2003; Lippus & Ross, 2011).

With this study we aim to address various issues concerning the production of the Estonian quantities in spontaneous speech. The first one concerns the duration of the stressed syllable rhyme in different syllable structures. The syllable rhyme in Estonian can be made up of different combinations of vowels and consonants. Here we will focus on words consisting of two open syllables (vocalic quantity) and those consisting of a closed stressed syllable followed by an open unstressed syllable (consonantal quantity). The realization of these two foot types has not been systematically compared in earlier studies. Based on the view that the three-way Estonian quantity contrast is not segmental but operates over a disyllabic foot, we hypothesize that the duration of the stressed syllable rhyme remains similar despite the syllable structure.

Secondly, we will investigate the stability of such secondary features of quantity as pitch and vowel quality in spontaneous speech. Even if some earlier studies have addressed the role of pitch cues in marking quantity in spontaneous speech (e.g. Asu et al., 2009; Krull, 1992, 1993a, 1993b), quantity-related variation of vowel quality in Estonian has received little attention so far. Our hypothesis is that the realization of Estonian quantity in spontaneous speech does not differ from the findings from lab speech, where the three quantities are differentiated by the temporal and pitch characteristics.

Thirdly, the paper aims at evaluating the impact of the different acoustic features on the quantity of the foot. We assume that the three-level opposition in a real-life situation should be generalizable over the numerous combinations of syllable structures that are possible in Estonian. Rather than creating a separate model for each syllable structure or accentuation type we will challenge the data with a single robust model, a multinomial logistic regression model. The same model was used in Lippus (2010) in order to characterize quantity as a function of segmental duration and vowel quality. A perception model by Traunmüller and Krull (2003) implies that the variability of the secondary features which separately are not significant can reveal their importance when combined in a model. We aim to test whether the same is true for production. Our hypothesis is that in the quantity model, the segmental durations are most significant, but in the opposition of Q1 vs. Q2 and Q3, vowel quality also plays a role, and in the opposition of Q1 and Q2 vs. Q3, an additional significant feature is the pitch contour.

2. Materials and method

The data come from the Tartu University Phonetic Corpus of Estonian Spontaneous Speech (http://www.murre.ut.ee/foneetikakorpus/), which currently comprises spontaneous dialogues and monologues from 37 speakers (30 h, ~215,000 words in total), and new material is regularly being added. The corpus is manually tagged using Praat (Boersma & Weenink, 2010), containing information about word, segment and syllable boundaries, quantity, and voice quality.

For present purposes, recorded materials from 21 speakers (8 women and 13 men) were used with a total duration of 18.5 h, including 3.5 h of monologues (recordings of lectures) and 15 h of spontaneous dialogues. Parts of this material were used in previous analyses dealing with related topics (Asu et al., 2009; Lippus, 2010). There was slightly more data from men (10 h) than from women (8.5 h). The dialogues were recorded in a purpose-built recording studio at the University of Tartu using high quality recording equipment. The monologues were recorded in various lecture halls using a headset microphone and field recording equipment. The speakers come from different places in Estonia, but all are speakers of Standard Estonian without noticeable dialectal traits. The age of the speakers varies between 21 and 58, the average age being 35.

The present study focuses on the realization of words with both vocalic and consonantal quantity. Therefore, disyllabic feet with two open syllables and feet where the first syllable is closed and the second syllable is open were extracted using a Praat script. Consonant clusters in a Q1 foot are not possible in Estonian, because in Q1 feet the first syllable is always short and open. Thus, a total of five different foot structures were analyzed: CVCV (Q1), CV:CV (Q2 V), and CV::CV (Q3 V) carrying the vocalic quantity opposition, and CVCCV (Q2 C) and CVC:CV (Q3 C) carrying the consonantal quantity opposition. In the case of vocalic quantity only words with monophthongs were used for the analysis, and for consonantal quantity only words containing a consonant cluster were used. In order to measure the F0 on the syllable boundary only such clusters where at least one of the two consonants was voiced were included.

It was shown in Asu et al. (2009) that the phrasal position of the word influences both the temporal and tonal characteristics of the quantity. In this paper, only words in phrase-medial position, i.e. not immediately preceded or followed by an intonational boundary, were included, because this position can be considered neutral with respect to boundary influences. Only words realized with the H*+L pitch accent, which is by far the most common intonational category in Estonian, were used in the analysis. There was too little material for other intonational categories to be included in the comparison. After eliminating tokens produced with creaky voice and whisper or those without pitch trace, the final data set consisted of 1716 words.

The duration of each segment was extracted from the annotated TextGrids using a Praat script. The F0 was automatically measured at four points in each word: the beginnings and ends of the syllable rhymes. All these measurements were subsequently checked and corrected by hand. An additional measurement was taken at the turning point (henceforth TP), where the F0 and the time were determined manually. The TP was defined as a pitch elbow (rather than a local F0 maximum), where the most noticeable change occurs in the general direction of the F0 contour in the stressed syllable from level or rising to falling, or in other words, the point with the highest rate of F0 change.

The words were grouped according to their accentuation as either accented or deaccented. Accentuation of each token was determined in the context where the word occurred. As shown before, in deaccented words, the pitch cue to the quantity is neutralized (Asu et al., 2009). Thus, in cases of deaccentuation there was normally no change in F0 and no TP could be located. The distribution of the data on the basis of accent condition and foot structure is presented in Table 1.

The F0 values in Hz were converted into semitones (st) in relation to 50, which is considered to be a good base-line for both male and female speakers. The semitone scale was used because it has been shown to be the most appropriate scale for comparing intonational data from several speakers (Nolan, 2003).

For analyzing the vowel quality, F1 and F2 values at the mid-point of V1 and V2 were extracted with a Praat script. All automatically measured formant values were manually checked, and converted to Bark using the formula by Traunmüller (1990). Subsequently, the vowel quality was normalized by calculating the Euclidean distance of each vowel from the so-called centroid vowel for each speaker (the procedure is discussed in detail in Harrington, 2010). The centroid vowel was taken as the mean of F1 and F2 values of each speaker's vowels [i, u, æ, ɑ]. The number of tokens for

Table 1
Number of analyzed words in the two accent conditions.

	Q1 CVCV	Q2 CV:CV	Q2 CVCCV	Q3 CV::CV	Q3 CVC:CV
Accented	530	154	95	138	138
Deaccented	419	54	74	52	62
Total	**949**	**208**	**169**	**190**	**200**

Table 2
Number of analyzed vowels in the first (S1) and the second (S2) syllable.

Gender	Quantity	S1									S2				
		æ	ø	ɤ	ɑ	e	i	o	u	y	ɑ	e	i	o	u
Female	Q1	32	0	19	54	47	27	28	23	3	99	47	25	0	62
	Q2 V	9	1	0	9	26	5	5	5	2	7	38	15	0	2
	Q2 C	0	0	2	21	9	19	2	2	2	4	10	25	0	18
	Q3 V	4	5	0	14	9	8	10	11	1	30	13	19	0	0
	Q3 C	19	0	3	8	7	17	2	6	1	31	8	20	0	4
Male	Q1	88	1	61	173	136	109	67	68	13	341	105	108	1	161
	Q2 V	13	1	1	19	31	32	28	19	2	22	95	26	0	3
	Q2 C	4	0	3	37	24	22	10	10	2	21	29	32	1	29
	Q3 V	10	20	2	30	21	15	18	10	2	69	17	40	0	2
	Q3 C	33	0	11	21	9	20	16	14	13	50	29	45	0	13

each vowel is presented in Table 2. Estonian has nine vowels [i, y, u, e, ø, ɤ, o, æ, ɑ] which can all occur in a primary stressed syllable. Only five vowels [i, u, e, o, ɑ] occur in non-initial syllables, and [o] can only be found in non-initial syllables in proper names and loan words.

The data was analyzed with the statistics package R (version 2.11.1). For the comparison of segment durations, pitch and formant values multifactor ANOVAs were used. The following dependent variables were tested: C1 duration, S1 rhyme duration, C2 duration, S2 rhyme duration, and S1/S2 ratio; in accented words, the pitch change from S1 onset to TP, from TP to S2 offset and the duration from S1 onset to TP; in deaccented words, the pitch change from S1 onset to S2 offset; the Euclidean distance of V1 and V2 from the center of the vowel space. The mean value of these acoustic measures was calculated for each speaker for each of the following factor conditions: Accent (accented and deaccented), Syllable structure (open and closed S1), and Quantity (Q1, Q2, Q3). The alpha level of the ANOVAs was set to 0.01, in order to reduce the potential overestimation of the significance level (Type I error) that might arise with a factorial ANOVA design. Post-hoc testing was carried out with Tukey's HSD.

3. Results and discussion

3.1. Duration

Table 3 presents the mean segment durations and standard deviations in Q1, Q2 (Q2 V—long vowel, Q2 C—short vowel plus syllable-final consonant), and Q3 (Q3 V, Q3 C) in accented and deaccented words. In the case of consonantal quantity, S1 rhyme is the sum of the duration of V1 and the following consonant. In all other cases, the syllable onset consists of a single syllable-initial consonant, and the rhyme is the vowel.

The mean S1 rhyme duration in Q1 was 57 ms in Q1, 116 ms in Q2, and 147 ms in Q3. The duration of S1 rhyme was significant both between Accent [$F(1, 167)=20.496$; $p<0.001$] and Quantity [$F(2, 167)=124.586$; $p<0.001$]. Surprisingly, the interaction of Accent and Quantity was not significant [$F(2, 167)=1.214$; $p=0.3$]; post-hoc testing showed significant differences between all quantity degrees within the accent conditions at $p<0.001$. The only exceptions were the combination of accented Q1 vs. deaccented Q1, and the combination of accented Q2 vs. deaccented Q3, which were not significantly different. From Fig. 1 it can be seen that S1 rhyme is longer in accented words than in deaccented words. It is the shortest in Q1, and the longest in Q3.

Given that the quantity is a foot-level phenomenon, we expected the syllable rhyme to have the same duration independent of its structure, within a given quantity degree. However, when S1 was a closed syllable, the overall rhyme duration was always significantly longer than when it was an open syllable (see the right panel of Fig. 1). Also, the short vowel in the consonantal Q2 and Q3 feet was longer than the short vowel in Q1 feet, but these differences were very small (up to 20 ms). The Syllable structure had a significant main effect [$F(1, 127)=10.436$; $p<0.01$], but again, none of the interactions with Quantity and Accent were significant.

There was some variation in the duration of the unstressed syllable rhyme (S2 rhyme = V2 as the data-set included only words with an open S2). There was a significant main effect of Accent [$F(1, 167)=17.064$; $p<0.001$] and Quantity [$F(2, 167)=27.532$; $p<0.001$], but the interaction between the two was not significant. V2 ranged from 70 to 50 ms, being significantly different in all quantity degrees in accented words ($p<0.01$), but not in deaccented words. The difference between the accentuation levels was significant only for Q1 ($p<0.01$).

The durational variation of the S2 rhyme is often ascribed to foot isochrony (Lehiste, 2003; Nolan & Asu, 2009), as the variation in S2 compensates for the variation in S1, or the quantity contrast is described as the syllable (rhyme) ratio. The left panel of Fig. 2 presents the foot

Table 3
Mean segment durations and standard deviations (in ms) in Q1, Q2 and Q3 accented and deaccented words.

	S1 onset		S1 rhyme		S2 onset		S2 rhyme		Total	
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
Q1										
Accented	63	8	61	9	56	10	87	16	267	31
Deaccented	56	7	53	11	55	10	68	11	232	26
Mean	60	8	57	11	56	10	79	16	251	34
Q2 V										
Accented	75	16	119	29	57	16	73	24	325	68
Deaccented	61	13	98	30	50	14	62	18	271	55
Mean	69	16	110	31	54	15	69	22	302	68
Q2 C										
Accented	72	17	132 (77+55)	30	49	9	61	13	307	56
Deaccented	64	11	109 (61+48)	27	44	10	51	17	262	52
Mean	68	15	122 (70+52)	31	47	10	56	16	286	58
Q3 V										
Accented	75	14	148	30	63	13	58	11	344	46
Deaccented	70	18	123	33	59	13	54	19	307	67
Mean	73	16	136	33	61	13	57	15	327	59
Q3 C										
Accented	63	9	171 (78+93)	32	57	12	53	10	338	57
Deaccented	61	12	143 (69+74)	21	55	15	51	9	303	42
Mean	62	10	158 (74+84)	31	56	13	52	9	321	52

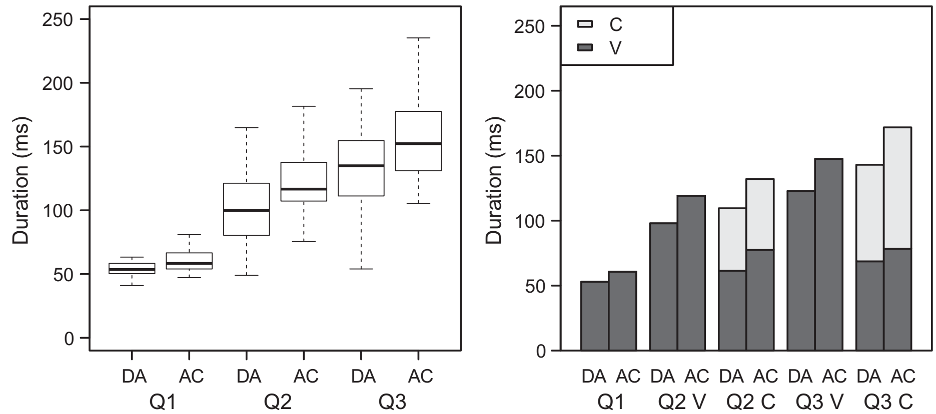

Fig. 1. The duration of S1 rhyme. Dark gray bars represent vowels, and light gray bars syllable final consonants. DA stands for deaccented, and AC for accented words. The left panel displays the distribution of S1 rhyme duration. The right panel shows the mean duration of the vowel (dark gray) and the offset consonant (light gray) of S1 in the words with vocalic quantity (Q2 V and Q3 V) and consonantal quantity (Q2 C and Q3 C).

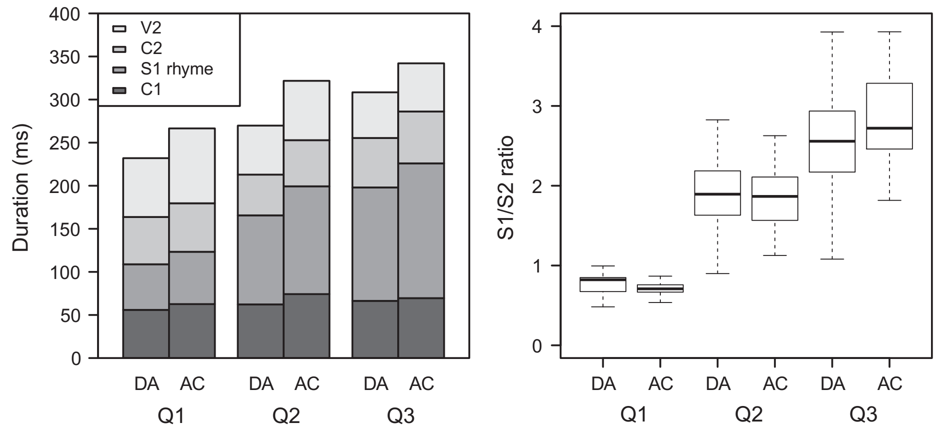

Fig. 2. Mean foot durations (left) and the S1/S2 duration ratio (right).

durations of the accented and deaccented words. The mean foot durations are roughly similar in all the quantities, although the difference between Q1 and Q2 is about 30 ms larger than between Q2 and Q3.

S1/S2 rhyme duration ratio is presented in the right panel of Fig. 2. There was a significant main effect of Quantity [$F(2, 167)=179.429$; $p<0.001$], and a non-significant effect of Accent and no interaction. Pairwise post-hoc testing showed significant differences between the quantity degrees ($p<0.001$ for all combinations), while the combinations of the same quantity degree with different accentuation conditions were always non-significant. Thus the syllable rhyme ratio is indeed a robust way for illustrating the three-way quantity opposition in Estonian.

The duration of syllable-initial consonants has been said to vary mainly due to speech rate and less due to the quantity degree (Eek & Meister, 2003; Lehiste, 1960). However, our results exhibit some quantity-related variation. The average duration of the initial consonant in the stressed syllable (C1) ranged from 56 to 75 ms and was significantly different for Accent [$F(1, 167)=12.569$; $p<0.001$] and Quantity [$F(2, 167)=6.625$; $p<0.01$], but there was no interaction [$F(2, 167)=1.82$; $p=0.165$]. C1 was shorter in Q1 and longer in Q2 and Q3; post-hoc testing revealed a difference between Q1 vs. Q2 and Q3.

The duration of the onset of the second syllable (C2) ranged from 44 to 63 ms, being shorter than the stressed syllable onset. The difference between levels of Accent failed to reach significance [$F(1, 167)=3.853$; $p=0.051$], but Quantity was fully significant [$F(2, 167)=7.259$; $p<0.001$]. Post-hoc testing showed that C2 was significantly shorter in Q2 than in other quantity degrees. There was no interaction of Accent and Quantity [$F(2, 167)=0.531$; $p=0.589$].

Finally, Table 4 presents the correlations between the duration of the segments within the foot. The correlations were significant between all segments. As expected, the correlation was strongest between the durations of syllable onsets and the S1 rhyme. Given that the quantity is typically described as the duration ratio of S1 and S2 rhymes, we would expect a strong negative correlation between the syllable rhymes. Therefore it is surprising that the correlation was rather weak. Furthermore, none of the correlations were strong enough to express all the variability of the other variable.

3.2. Tonal characteristics

Fig. 3 presents the mean pitch contours for accented and deaccented words in different quantity degrees. The pitch contours of the deaccented words are shown in the left column. It can be seen that in such words the pitch fell slightly from the beginning of each syllable but the overall pitch change within the foot remained in the range of about one semitone. In deaccented words there were no significant differences in F0 movement from S1 onset to S2 offset between the three quantity degrees [$F(2, 76)=1.285$; $p=0.282$].

Table 4
The Pearson product–moment correlations between the durations of segments.

	C1	S1 rhyme	C2
S1 rhyme	0.404*		
C2	−0.047***	0.532*	
S2 rhyme	0.191***	−0.066***	−0.289*

Signif. codes: *** stands for $p<0.001$ and * for $p<0.05$.

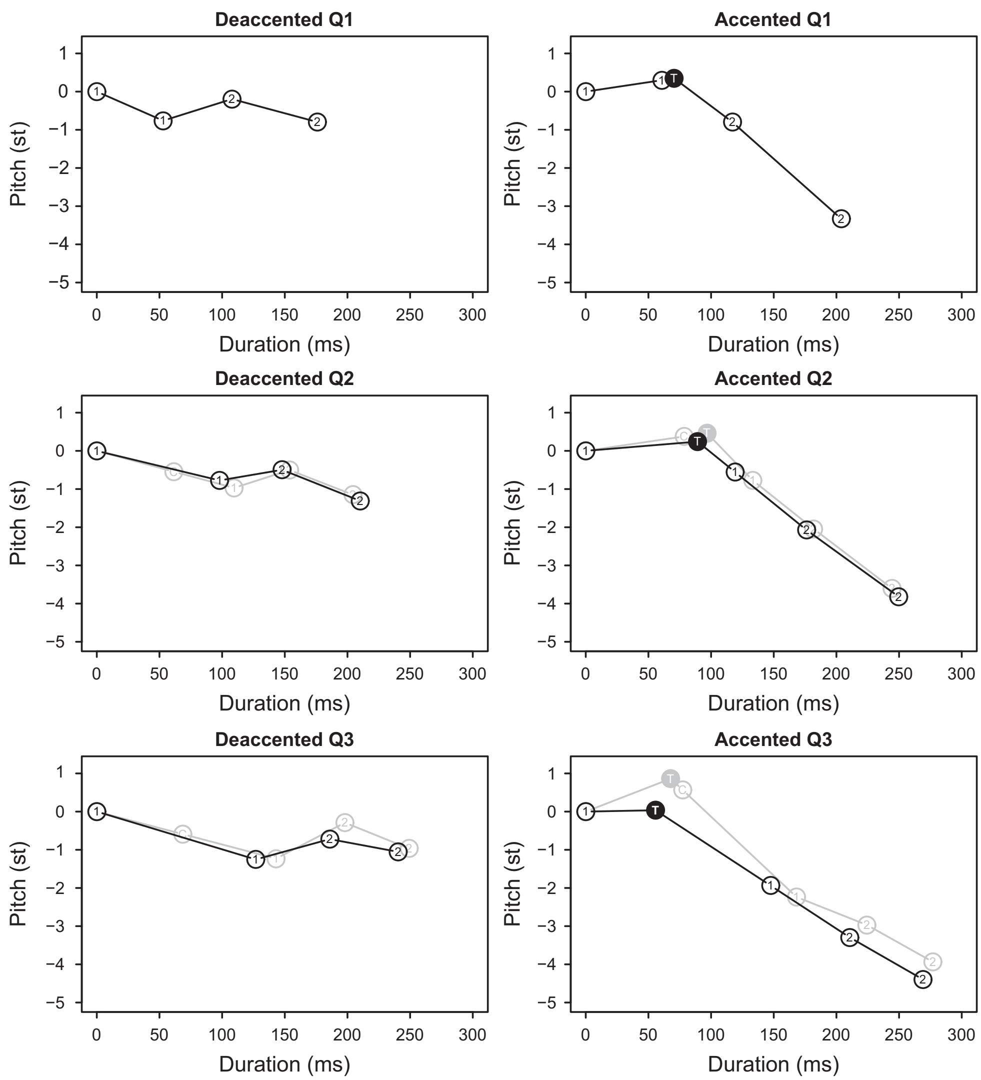

Fig. 3. Pitch contours in accented and deaccented words in the three quantities. The *Y*-axis shows the pitch movement normalized relative to the beginning of the word (st), and the *X*-axis the duration of the word (ms). Black lines represent the words with vocalic quantity (open S1), and gray lines those with consonantal quantity (closed S1). Circles marked with number 1 represent measurement points at the beginning and end of S1 rhyme, and those marked with 2 show the measurement points at the beginning and end of S2 rhyme. The turning point is marked with the letter T and the beginning of the coda consonant in consonantal quantity words with the letter C.

The pitch contours of the accented words are given in the right column of Fig. 3. The pitch change from S1 onset to TP was less than a semitone. There was no significant effect of Quantity [$F(2, 91)=0.044$; $p=0.957$]. Likewise, Quantity was not significant for the pitch change from TP to S2 offset, where the pitch fell on average about 4 st in all cases [$F(2, 91)=2.838$; $p=0.064$].

As can be seen from Fig. 3, there was no significant effect of Syllable structure to the alignment of TP in relation to the S1 boundary. If however, the alignment of the TP is viewed with respect to the syllable structure of S1, it emerges that both in Q2 and Q3, the TP is aligned roughly with the end of the vowel either just before (Q2) or after (Q3) the coda consonant. A similar result was obtained by Plüschke (2011) on the basis of controlled read sentences.

The left panel of Fig. 4 presents the location of the TP as an absolute duration in relation to the beginning of V1. The distance from S1 onset to the TP was 71 ms in Q1, 92 ms in Q2, and 61 ms in Q3. There was a significant effect of Quantity [$F(2, 91)=13.546$; $p<0.001$], but the post-hoc testing showed no significant difference between Q1 vs. Q3. The picture becomes clearer in the right panel of Fig. 4, where the location of the TP is shown in relation to the syllable boundary within the foot expressed as a relative duration. The TP was located in the middle of the

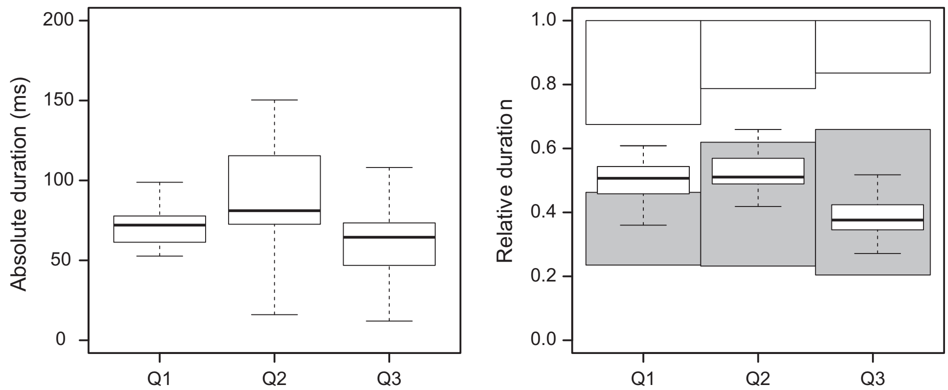

Fig. 4. The distance of the TP from the beginning of V1 in ms (left) and the location of the TP relative to the foot duration (right). The stressed syllable rhyme is marked with a gray box and the unstressed syllable rhyme with a white box in the background. The 0 point marks the onset of C1.

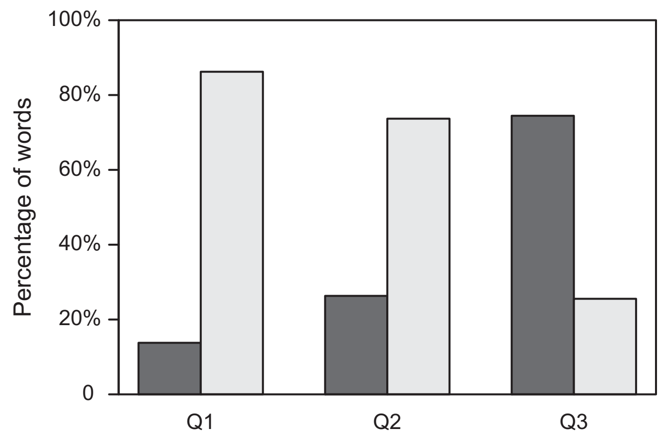

Fig. 5. Percentage of words where the TP is located within the first half of S1 (dark gray bars) and where it is located later (light gray bars).

disyllabic foot in Q1 and Q2, but earlier in Q3 [$F(2, 91)=29.692$; $p<0.001$]. Post-hoc tests showed no significant difference between Q1 and Q2, but Q3 was different from both Q1 and Q2 at $p<0.001$. In Q1, the TP was at the beginning of the second syllable, in Q2, in the second half of the first syllable and in Q3, in the first half of the first syllable. There was some overlap between the quantity degrees in the TP location. Fig. 5 shows the percentage of words where the TP is located in the first half of the S1 rhyme (the cut-off point was placed at 55%). We can see a relatively neat pattern emerging: the TP was in the first half of the S1 rhyme in about 75% of the Q3 words as compared to only about 25% of the Q1 and Q2 words.

3.3. Vowel quality

The vowel formant values were converted into Bark and normalized to Euclidean distance from the central point of the vowel space of every individual speaker. This was done in order to eliminate the phonological distinctions between the vowel categories and individual differences between the speakers, while providing a rough measure of centralization that can be expressed by intervals on a psycho-acoustic scale.

In Fig. 6, the vowels are plotted in the F1–F2 space. The vowels /y, ø, ɤ/ and the unstressed vowel /o/ were not included because there were fewer than five observations of each vowel in the Q2 and Q3 feet. The analysis showed no quality difference between the stressed vowels of Q2 and Q3 feet. Therefore, the vowels are presented in three groups: short vowels in Q1 feet, short vowels in Q2 and Q3 feet with consonantal quantity, and long vowels in Q2 and Q3 feet with vocalic quantity.

The quality of the stressed vowels showed a significant main effect of Accent [$F(1, 163)=12.053$; $p<0.001$] and Quantity [$F(2, 163)=22.536$; $p<0.001$], but not of Syllable structure [$F(1, 163)=4.356$; $p=0.031$]. None of the interactions were significant. Post-hoc testing shed more light on the patterns. On average, the stressed vowels in deaccented words were positioned at 1.9 Bark from the center of the vowel space while the accented words were more peripheral, at 2.2 Barks from the center point. However, the pairwise difference between the same quantity degrees in different accent conditions was not significant. There was a significant difference between Q1 (1.6 Bark from the center of the vowel space) vs. vocalic Q2 and Q3 (2.5 and 2.3 Bark respectively) at $p<0.001$, and between vocalic Q2 and Q3 vs. consonantal Q2 and Q3 (both 1.9 Bark) at $p<0.01$. At the same time, there was no significant difference between Q2 and Q3 within foot structure and between Q1 vs. consonantal Q2 and Q3.

Thus, the short stressed vowels were closer to the center in Q1, while the long and overlong vowels in Q2 and Q3 were more peripheral. The difference between long and overlong vowels was, however, very small. Moreover, the short vowels in Q2 and Q3 with consonantal quantity were nearly as central as the short vowels in Q1 feet. The variation in vowel quality seems to be related to vowel duration and not so much to the three-way foot-level quantity distinction. There was also a significant correlation ($r=0.469$) between the V1 duration and its distance from the center of the vowel space.

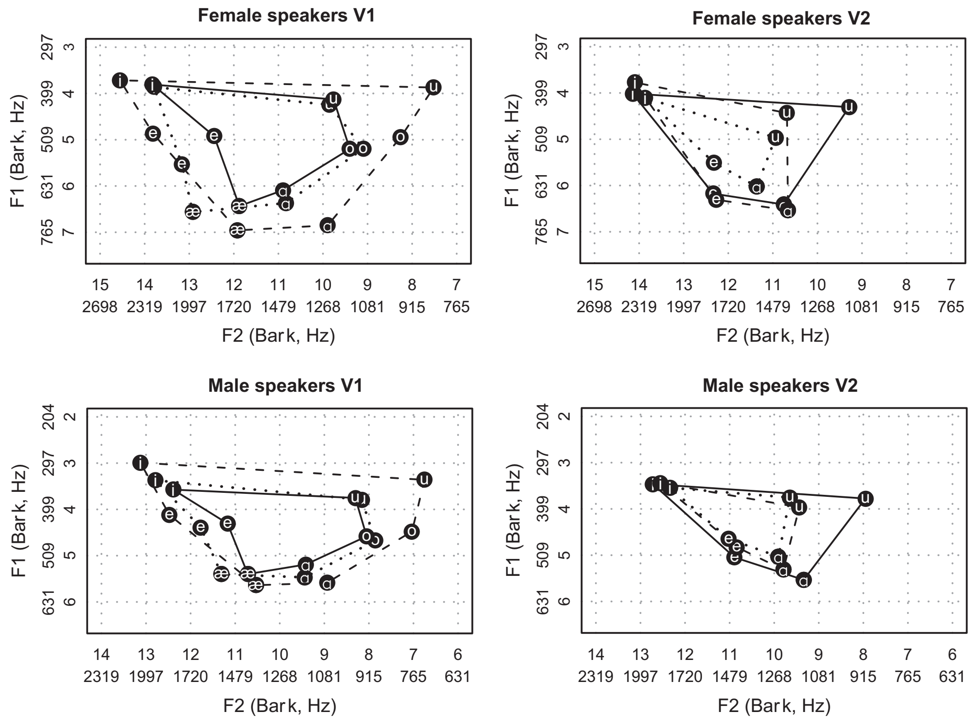

Fig. 6. Stressed and unstressed vowels in the F1–F2 space. The panels on the left show the stressed vowels: short vowels from Q1 words (solid line) and Q2 and Q3 consonantal quantity words (dotted line), and long vowels from Q2 and Q3 vocalic quantity words (dashed line). The panels on the right show the unstressed vowels from Q1 words (solid line), Q2 words (dashed line), and Q3 words (dotted line).

Unstressed vowels showed more variation in general, and although in Fig. 6 one can see a pattern where the vowels in Q2 and Q3 are more centralized than in Q1, there was no significant effect of Accent [$F(1, 163)=3.077$; $p=0.081$] or Quantity [$F(2, 163)=1.614$; $p=0.202$]. Also, the correlation between vowel duration and centralization was not significant. The unstressed vowels in general are more centralized than the stressed vowels (the distance from the center of the vowel space was 1.7 Bark). It is worth noting that the location of unstressed /e/ was not centralized but lowered, being similar to that of /æ/ in a stressed syllable.

3.4. Multinomial logistic regression model

In order to test all the above measured acoustic features simultaneously, a multinomial logistic regression model was used. The three-level quantity opposition was selected as the dependent variable with Q2 as the reference level. The independent variables were selected so that the model could be built to apply as universally as possible and accommodate each accent condition and foot structure. The independent variables were C1 duration, S1 rhyme duration, C2 duration, and S2 rhyme duration (measured in ms), S1 F0 range (st), V1 Euclidean distance and V2 Euclidean distance (from the centroid; Bark).

In the model, the location of the TP (which was above expressed as the duration from S1 onset to TP) was replaced with a measure of the change of F0 from the onset to the offset of S1. This was done because there was no measurable TP in the deaccented words, as there was no significant pitch movement. In accented words, the total F0 range within the foot was the same in all the quantity degrees, but the TP location determined how dynamic the pitch was in the stressed syllable. Consequently the S1 F0 range (i.e. the F0 range within the stressed syllable) reflects the TP location: the earlier the TP is located, the more the pitch falls by the end of the syllable.

As expected, the significant factors for the model were the segmental durations and vowel quality in the case of Q1 vs. Q2 opposition, and the segmental durations and pitch movement in the case of Q3 vs. Q2 opposition. The model shows the probability of the word being in one of the three quantity degrees as a function of variation in the independent variables, which enables us to evaluate the impact of each independent variable by exponentiating the coefficient values. The last column in Table 5, exp(*b*), expresses the relative change of the probability of the quantity having a level other than the reference level if the value of that independent variable was raised by one unit.[2] By comparing the impact of the one unit change to the range of that variable in the real data we can judge the relative significance of each measure.

C1 duration gives a −1% impact in case of 1 ms duration change for both Q1 and Q3. C2 duration onset has a 3–4% impact in the same conditions. The effect is significant but rather small, in particular if we consider that the mean duration of the syllable onset consonants varies less than 10 ms between the quantity degrees. Yet, the syllable onsets were systematically shorter in Q2 than in Q1 and Q3. We can agree with Traunmüller and Krull (2003) who note that even if most of the variation in the duration of syllable onsets can be accounted for by the local speaking rate, there is a weak effect of quantity.

Among the temporal features, the S1 rhyme duration has the greatest impact. In the case of Q1 vs. Q2, a 1 ms longer duration results in a −12% impact on the probability of Q1, while the mean S1 rhyme is 64 ms longer in Q2 than in Q1. In the case of Q3 vs. Q2 the impact of S1 rhyme duration is smaller, only 3%, while the mean S1 rhyme is 31 ms longer in Q3 than in Q2. Also, the impact of the S2 rhyme duration is

[2] The value of exp(*b*) is the coefficient by which the probability of reference level vs. alternative should be multiplied. Thus, a value of 1.00 has no impact, while <1 values lower the probability of alternative decision and >1 values raise it (e.g. a value of 0.95 would give a −5% impact while 1.05 results in +5%).

Table 5
Multinomial logistic regression analysis of the quantity in the foot (with Q2 as the reference level).

	b	S.E.	*t*	Pr(>|*t*|)	exp(*b*)
Q1 vs. Q2					
Intercept	5.329	0.695	7.662	0.000***	
C1 duration	−0.015	0.006	−2.588	0.010**	0.985
S1rhyme duration	−0.13	0.009	−13.871	0.000***	0.878
C2 duration	0.0353	0.008	4.387	0.000***	1.036
S2 rhyme duration	0.081	0.008	10.262	0.000***	1.084
S1 F0 range	−0.088	0.133	−0.659	0.510	0.916
V1 Euc. distance	−0.800	0.146	−5.477	0.000***	0.449
V2 Euc. distance	0.672	0.149	4.500	0.000***	1.958
Q3 vs. Q2					
Intercept	−2.490	0.440	−5.658	0.000***	
C1 duration	−0.010	0.004	−2.738	0.006**	0.990
S1rhyme duration	0.025	0.003	9.788	0.000***	1.026
C2 duration	0.028	0.005	5.637	0.000***	1.028
S2 rhyme duration	−0.040	0.005	−8.445	0.000***	0.961
S1 F0 range	0.404	0.066	6.114	0.000***	1.497
V1 Euc. distance	0.057	0.089	0.638	0.524	1.058
V2 Euc. distance	−0.076	0.103	−0.739	0.460	0.927

*** stands for $p<0.001$ and ** $p<0.01$.
Log-Likelihood: −589.81.
McFadden $R^2=0.653$.
Likelihood ratio test: $\chi^2=2219$ ($p<0.001$).

Table 6
The probability of the quantity predicted by the model using the mean values.

	p_{Q1}	p_{Q2}	p_{Q3}
Accented Q1	0.995	0.005	0.000
Accented Q2	0.003	0.666	0.331
Accented Q3	0.000	0.213	0.787
Deaccented Q1	0.993	0.006	0.000
Deaccented Q2	0.024	0.724	0.252
Deaccented Q3	0.000	0.463	0.537

greater in the case of Q1 vs. Q2 (+8%) than in the case of Q3 vs. Q2 (−4%), while the range is about the same in both cases: S2 rhyme is about 18 ms longer in Q1 than in Q2 and about 13 ms shorter in Q3 than in Q2.

In the model, the pitch contour is not a significant feature in the opposition of Q1 vs. Q2. Even though the TP is outside the stressed syllable in Q1 and towards the end of the stressed syllable in Q2, the mean difference between Q1 and Q2 in the S1 F0 range is about 0.5 st. In the case of Q3 vs. Q2, on the other hand, the impact of S1 F0 range is considerably larger. Due to an earlier TP in Q3, the pitch changes in S1 about 1.2 st more than in Q2, resulting in a 50% impact.

In the case of Q1 vs. Q2 there is a significant impact on vowel quality. The stressed vowel in Q1 is more central by 0.8 Bark than in Q2. In the case of consonantal quantity, the difference between the three quantities is negligible, which suggests that vowel quality is more closely connected with duration and not so much with quantity. Still, vowel quality is one additional feature which helps to differentiate between the quantities. There is no significant variation in vowel quality between Q2 and Q3.

Somewhat unexpectedly, in the model, the V2 Euclidean distance was significant between Q1 and Q2 but not between Q2 and Q3. The variation is rather small in both cases (the mean V2 is more peripheral by about 0.2 Bark in Q1 than in Q2 and in Q2 than in Q3) and not significant when viewed separately but in the model there is a weak effect.

Finally, the goodness of fit of the model can be evaluated by using it for predicting the quantity of words on the basis of mean values. The mean values of the acoustic measures that were used as independent variables in the model were calculated for six subsets of the data: accented Q1, accented Q2, accented Q3, deaccented Q1, deaccented Q2, and deaccented Q3. The probability of the subsets predicted to be Q1, Q2, or Q3 is presented in Table 6. The table shows that the model was able to predict Q1 very well in both accent conditions (the subset with the mean values of Q1 was predicted to be Q1 with the likelihood of 99%). For Q2 and Q3 subsets the prediction level was around 70% in the case of accented words, where the F0 range was an additional variable. In deaccented words, the prediction level was significantly lower for Q3.

4. General discussion

The first hypothesis of this paper concerned the duration of the stressed syllable rhyme, which is phonetically the main carrier of the quantity opposition. We hypothesized that the syllable rhyme has the same duration independent of its structure so that it could participate in the ternary length opposition of the foot. The results showed, however, that the rhyme duration of the closed syllable was longer than that of the open syllable. This small difference between the two foot structures could be explained by the effect of more complex linguistic content, as it is known that the production of units consisting of multiple articulatory movements takes longer than equivalent units that consist of a single movement (Lehiste, 1970).

Table 7
S1/S2 ratios as compared to earlier data.

	Q1	Q2 V/Q2 C	Q3 V/Q3 C
Lehiste (1960)	0.7	1.5	2.0
Liiv (1961)	0.7	1.6	2.6
Eek (1974)	0.7	2.0/1.6–1.9	3.9/2.8–3.0
Eek and Meister (1997)	0.7	1.3	2.8
Krull (1991, 1992)	0.5–0.7	1.2–2.1	2.2–2.9
Krull (1993a)	0.7	1.7	2.6
Krull (1993b)	0.7	1.8	3.2
Krull (1997)	0.7	1.7	2.6
Lippus and Ross (2011)	0.8	2.0/2.0	2.9/3.6
Present study	0.8	1.8/2.3	2.6/3.4

It has been established in earlier studies on Estonian (e.g. Meister & Meister, 2011) as well as for other languages with a three-way length distinction (e.g. Dinka, see Remijsen & Gilley, 2008) that the quantity difference manifested by the stressed syllable rhyme is not equidistantly distributed. In our data, the duration of the S1 rhyme in Q1 was about half of the duration of the S1 rhyme in Q2, i.e. the long vowel was twice as long as the short vowel. The overlong vowel was not three times longer than the short vowel, but less than two and a half times longer. The Q3 foot was only about 20% longer than the Q2 foot.

It has been shown that the duration of the stressed syllable alone is not sufficient for distinguishing between the quantity degrees (Eek & Meister, 2003). Thus, it is the duration of the second syllable that comes into play at this point. It is worth stressing that although, on the whole, the influence of accentuation is apparent in all segmental durations, the duration of V2 in Q3 does not match this pattern, i.e. V2 in Q3 is kept extra short. At the same time the variation in V2 duration in the three quantities is very small.

The Estonian quantity system based on feet has usually been interpreted using the concept of isochrony (e.g. Lehiste, 2003). Partly connected to this are the notions of half-long, short, or extra-short used to describe the duration of unstressed vowels in Q1, Q2 or Q3 disyllabic feet, respectively. Unstressed vowels do not carry a phonological length opposition. The concept of isochrony has been hotly debated (see e.g. Arvaniti, 2009), giving the impression that there is no unified understanding as to its essence. In Estonian, isochrony is to be understood as a tendency rather than a statistically significant equal duration of feet. It is hard to say whether this inverse co-variation of the stressed and unstressed syllable duration, which has been described as isochrony, is due to an attempt to achieve feet of equal duration, or whether instead the shorter duration of the unstressed syllable is a way to enhance the perceived duration of the stressed syllable.

Traditionally, the Estonian quantity opposition has been expressed as the ratio of S1/S2. Table 7 presents syllable ratios calculated on the basis of our data as compared to the results from several earlier studies. Following the common procedure suggested by Lehiste (1960), the onset consonant is usually left out of the calculation of syllable duration, because this consonant is always short and does not carry phonological length. Perhaps not surprisingly, our results are most similar to those by Krull (1993a, 1993b, 1997), which were also based on words extracted from spontaneous speech. However, broadly speaking they are also comparable to other earlier studies.

Alternatively, Eek and Meister (2003, 2004) have proposed an equation to describe the opposition of Q2 and Q3, where the durations of the nucleus and coda of S1 are compared with the nucleus of S2. Following this, Q2 can be expressed as S1 nucleus $\geq$ S1 coda $\leq$ S2 nucleus, and Q3 as S1 nucleus $<$ S1 coda $>$ S2 nucleus. Our data match this description, although it is doubtful whether this comparison would be more plausible as a perceptual account. Furthermore, it is possible to apply it objectively only in feet with consonantal quantity.

According to our second hypothesis, the secondary cues to the quantity such as pitch and vowel quality are realized similarly in spontaneous speech and in read speech. Our results confirmed this hypothesis. The analysis showed that on an absolute scale, the pitch alignment in words with the H*+L pitch accent was rather similar in all the quantity degrees. There was a small (though significant) difference between Q2 vs. Q1 and Q3. This is not in line with the common description of quantity-related pitch variation in Estonian according to which the difference in TP distance is between Q1 and Q2 vs. Q3. The pitch movement which significantly discriminates between Q2 vs. Q3 should be viewed in relative terms in relation to the syllable boundary or the duration of the foot.

The third aim of the paper was to evaluate the importance of different quantity characteristics within one model. In order to test all the acoustic features simultaneously we used a multinomial logistic regression model. Unlike the perception model by Traunmüller and Krull (2003) our model also included measures of pitch and vowel quality. The model showed that all the segment durations play a significant part in the three-way quantity opposition. Yet the S1 rhyme did have the highest weight among the temporal features. The pitch was significant for the opposition of Q2 vs. Q3 and the vowel quality for the opposition of Q1 vs. Q2.

Although the model showed vowel quality as a significant feature in the opposition of Q1 and Q2, the separate analysis of vowel quality revealed no categorical difference. The short stressed vowels in Q1 were most centralized and the long vowels in Q2 and Q3 most peripheral. The short vowels in consonantal Q2 and Q3 were centralized like the short vowels of Q1 feet and unlike the long vowels in vocalic Q2 and Q3.

The vowels of the unstressed syllable were reduced as compared to the vowels in the stressed syllable. While the quantity-related variation of vowel quality (as measured in Euclidean distance from the center of the vowel space) was not significant when viewed separately, the model found this variation between Q1 and Q2 to be significant. Additionally, Eek and Meister (1998) report a tendency for the unstressed /e/ in Q1 feet to be perceived as /æ/. The data in the present study showed that the /e/ in the second syllable had changed into an open front vowel in all quantities. Therefore, the lowering of a close-mid-vowel further away from the existing close vowel towards an empty corner in the vowel space may be viewed as the pursuit of maximal contrast (Liljencrants & Lindblom, 1972), which consequently in some sense can be interpreted as reduction (Crosswhite, 2004). Quality contrasts are more vulnerable in unstressed syllables because of their reduced prominence; hence the vowel space may be expanded to compensate for the potential loss in salience of the contrast.

The model would of course benefit from additionally taking into account the intonation and the phrasal position (as suggested in Asu et al., 2009). Nevertheless, even such a robust model is able to predict the three-way quantity on a level higher than chance. As compared to the model tested in Lippus (2010) the present model gave a better performance with regard to accented words, but was not considerably better in the case of deaccented words. The model in Lippus (2010) did not include pitch as a factor, and predicted Q2 vs. Q3 nearly at a chance level. The inclusion of pitch as a factor raised the model's ability of predicting Q2 vs. Q3 to about 70% in the case of accented words.

5. Conclusions

This paper has investigated the realization of duration, pitch and vowel quality as quantity characteristics in spontaneous Estonian. Broadly, the results show that the realization of Estonian quantities in spontaneous speech is in line with earlier descriptions of the quantities in read speech.

The rhyme components of the stressed syllable depend on whether the foot is of vocalic or consonantal quantity, but other than that, from the point of view of quantity, the two patterns are not distinct.

The durational characteristics that are important for the quantity are the durations of the syllable rhymes. There is some variation in the duration of syllable onsets, but their effect on the quantity is negligible. The duration of the stressed and unstressed syllables varies inversely, and thus a slight tendency to isochrony becomes apparent. It is difficult to say whether the shortening of the unstressed syllable caused by the lengthening of the stressed syllable signals an attempt to achieve foot isochrony or whether, rather, the shortness of the unstressed syllable promotes the impression of greater length of the stressed syllable.

F0 is an important characteristic in distinguishing Q2 and Q3, as has also been shown on the basis of controlled speech. In absolute terms the fall in F0 occurred at more or less the same location in all the quantities, but due to variation in the duration of the stressed syllable, the TP was located close to the syllable boundary in Q1 and Q2, and in the first half of the stressed syllable in Q3.

Vowel quality varied in step with the duration: the quality of short vowels was more centralized and that of long vowels more peripheral. This is above all due to the duration of the vowel rather than the quantity. Short stressed vowels in Q2 and Q3 feet with consonantal quantity were shown to be centralized similarly to short stressed vowels in Q1. Likewise, no significant difference in the quality of long vowels in Q2 and Q3 was observed. Similarly, since pitch as a quantity characteristic can be closely attributable to the duration, while at the same time being a perceptually significant feature, vowel quality can also be viewed as an additional feature of the quantity. Additionally, unstressed vowels in non-initial syllables reduce in two directions: on the whole they are more centralized, but /e/ was lowered to [æ] in order to provide maximal separation of the four vowels that occur in the Estonian unstressed syllable.

Acknowledgments

We would like to thank three anonymous reviewers and the Associate Editor Taehong Cho whose comments and corrections greatly improved the paper. We are also grateful to Stefan Werner and Karl Pajusalu for their comments on a draft of this paper, and Virve Vihman for proof-reading the English. Research presented in this paper was partly supported by the Estonian Science Foundation Grant no. 7904, the National Program for the Estonian Language Technology Project EKT4, and the target-financed Theme no. SF0180076s08 of the Estonian Ministry of Education and Research.

References

Andersen, T. (1987). The phonemic system of Agar Dinka. *Journal of African Languages and Linguistics*, *9*, 1–27.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*(1–2), 46–63.

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

Asu, E. L., & Nolan, F. (1999). The effect of intonation on pitch cues to the Estonian quantity contrast. In *Proceedings of the 14th international congress of phonetic sciences* (Vol. 3, pp. 1873–1876). San Francisco, USA, 1–7 August 1999.

Asu, E. L., & Nolan, F. (2007). The analysis of low accentuation in Estonian. *Language and Speech*, *50*(4), 567–588.

Asu, E. L., & Teras, P. (2009). Illustrations of the IPA: Estonian. *Journal of the International Phonetic Association*, *39*(3), 367–372.

Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer* (Version 5.2) [Computer program]. Retrieved October 2010, from 〈http://www.praat.org/〉.

Bremer, O. (1927). Der Schleifton im Nordniedersächsischen. *Niederdeutsches Jahrbuch*, *53*, 1–32.

Bye, P. (1997). A generative perspective on 'overlength' in Estonian and Saami. In: I. Lehiste, & J. Ross (Eds.), *Estonian prosody: Papers from a symposium* (pp. 36–68). Tallinn: Institute of Estonian Language.

Bye, P., Sagulin, E., & Toivonen, I. (2009). Phonetic duration, phonological quantity and prosodic structure in Inari Saami. *Phonetica*, *66*(4), 199–221.

Crosswhite, K. M. (2004). Vowel reduction. In: B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 191–231). Cambridge, New York: Cambridge University Press.

Eek, A. (1974). Observations on the duration of some word structures: I. *Estonian Papers in Phonetics*, 18–31.

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

Eek, A., & Meister, E. (1997). Simple perception experiments on Estonian word prosody: Foot structure vs segmental quantity. In: I. Lehiste, & J. Ross (Eds.), *Estonian prosody: Papers from a symposium* (pp. 71–99). Tallinn: Institute of Estonian Language.

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*, *34*(3), 226–233.

Eek, A., & Meister, E. (2003). Foneetilisi katseid ja arutlusi kvantiteedi alalt (I). Häälikukestusi muutvad kontekstid ja välde. *Keel ja Kirjandus*, *46*(11–12), 815–837 (904–918).

Eek, A., & Meister, E. (2004). Foneetilisi katseid ja arutlusi kvantiteedi alalt (II). Takt, silp ja välde. *Keel ja Kirjandus*, *47*(4–5), 251–271 (336–357).

Fox, R. A., & Lehiste, I. (1987). Discrimination of duration ratios by native English and Estonian listeners. *Journal of Phonetics*, *15*(4), 349–363.

Fox, R. A., & Lehiste, I. (1989). Discrimination of duration ratios in bisyllabic tokens by native English and Estonian listeners. *Journal of Phonetics*, *17*(3), 167–174.

Harrington, J. (2010). *Phonetic analysis of speech corpora*. Malden, MA: Wiley-Blackwell.

Hoogshagen, S. (1959). Three contrastive vowel lengths in Mixe. *Zeitschrift für Phonetik und Allgemeine Sprachwissenshcaft*, *12*(1–4), 111–115.

Järvikivi, J., Vainio, M., & Aalto, D. (2010). Real-time correlates of phonological quantity reveal unity of tonal and non-tonal languages. *PLoS ONE*, *5*(9), e12603, http://dx.doi.org/10.1371/journal.pone.0012603.

Krull, D. (1991). Stability in some Estonian duration relations. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS)*, *13*, 57–60.

Krull, D. (1992). Temporal and tonal correlates to quantity in Estonian. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS)*, *15*, 17–36.

Krull, D. (1993a). Temporal and tonal correlates to quantity in Estonian spontaneous speech: Some preliminary results. *Papers from the Seventh Swedish phonetics conference held in Uppsala*, May 12–24, 1993 (pp. 89–93). RUUL 23.

Krull, D. (1993b). Word-prosodic features in Estonian conversational speech: Some preliminary results. In *Proceedings of the PERILUS XVII, Experiments in speech processes* (pp. 45–54). Department of Linguistics, Stockholm University.

Krull, D. (1997). Prepausal lengthening in Estonian: Evidence from conversational speech. In: I. Lehiste, & J. Ross (Eds.), *Estonian prosody: Papers from a symposium* (pp. 136–148). Tallinn: Institute of Estonian Language.

Krull, D. (1998). Perception of Estonian word prosody. A study of words extracted from conversational speech. *Linguistica Uralica*, *34*(3), 167–171.
Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. In: T. A. Sebeok (Ed.), *American studies in Uralic linguistics, Uralic and Altaic series*, Vol. 1 (pp. 21–82). Bloomington: Indiana University Publications.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: M.I.T. Press.
Lehiste, I. (1975). Experiments with synthetic speech concerning quantity in Estonian. In V. Hallap (Ed.), *Congressus Tertius Internationalis Fenno-Ugristarum Tallinnae habitus 17.–23. VIII 1970. Pars I Acta linguistica* (pp. 254–269). Tallinn: Valgus.
Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, *4*, 113–117.
Lehiste, I. (1997). Search for phonetic correlates in Estonian prosody. In: I. Lehiste, & J. Ross (Eds.), *Estonian prosody: Papers from a symposium* (pp. 11–35). Tallinn: Institute of Estonian Language.
Lehiste, I. (2003). Prosodic change in progress: From quantity language to accent language. In: P. Fikkert, & H. Jacobs (Eds.), *Development in prosodic systems* (pp. 47–65). Berlin, New York: Mouton de Gruyter.
Lehiste, I., & Danforth, D. G. (1977). Foneettisten vihjeiden hierarkia viron kvantiteetin havaitsemisessa. *Virittäjä*, *81*(4), 404–411.
Lehiste, I., Teras, P., Ernštreits, V., Lippus, P., Pajusalu, K., Tuisk, T., et al. (2008). *Livonian prosody.* Suomalais-Ugrilaisen Seuran Toimituksia 255. Mémoires de la Société Finno-Ougrienne 255. Helsinki.
Lehnert-LeHouillier, H. (2010). A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics*, *38*(3), 472–482.
Liiv, G. (1961). Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid. *Keel ja Kirjandus*, *4*(7–8), 412–424 (480–490).
Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), 839–862.
Lippus, P. (2010). Variation in vowel quality as a feature of Estonian quantity. In M. Hasegawa-Johnson, A. Bradlow, J. Cole, K. Livescu, J. Pierhumbert, & C. Shin (Eds.), *Speech prosody 2010*. Chicago, USA. 〈http://aune.lpl.univ-aix.fr/~sprosig/sp2010/papers/100877.pdf〉.
Lippus, P., & Pajusalu, K. (2009). Regional variation in the perception of Estonian quantity. In: M. Vainio, R. Aulanko, & O. Aaltonen (Eds.), *Nordic prosody. Proceedings of the Xth conference, Helsinki 2008* (pp. 151–157). Frankfurt: Peter Lang.
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.
Lippus, P., Pajusalu, K., & Allik, J. (2011). The role of pitch cue in the perception of the Estonian long quantity. In: S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 231–242). Dordrecht: Springer.
Lippus, P., & Ross, J. (2011). Has Estonian quantity system changed in a century? Comparison of historical and contemporary data. In: W.-S. Lee, & E. Zee (Eds.), *Proceedings of the 17th international congress of phonetic sciences* (pp. 1262–1265). Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong.
Markus, E. (2011). The phonetics and phonology of a disyllabic foot in Soikkola Ingrian. *Lingustica Uralica*, *47*(2), 103–119.
McRobbie-Utasi, Z. (2007). The instability of systems with ternary length distinctions. In I. Toivonen, & D. Nelson (Eds.), *The Scolt Saami evidence—Saami linguistics* (pp. 167–206). Current Issues in Linguistic Theory 288. Amsterdam: John Benjamins Publishing Company.
Meister, E., & Werner, S. (2009). Duration affects vowel perception in Estonian and Finnish. *Linguistica Uralica*, *45*(3), 161–177.
Meister, E., Werner, S., & Meister, L. (2011). Short vs. long category perception affected by vowel quality. In: W.-S. Lee, & E. Zee (Eds.), *Proceedings of the 17th international congress of phonetic sciences* (pp. 1362–1365). Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong.
Meister, L., & Meister, E. (2011). Perception of the short vs. long phonological category in Estonian by native and non-native listeners. *Journal of Phonetics*, *39*(2), 212–224.
Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. In: M. J. Solé, D. Recasens, & J. Romero (Eds.). *Proceedings of the 15th international congress of phonetic sciences* (pp. 771–774). Barcelona, Spain, 3–9 August 2003.
Nolan, F., & Asu, E. L. (2009). The pairwise variability index and coexisting rhythms in language. *Phonetica*, *66*(1–2), 64–77.
Plüschke, M. (2011). Peak alignment in falling accents in Estonian. In: W.-S. Lee, & E. Zee (Eds.), *Proceedings of the 17th international congress of phonetic sciences* (pp. 1614–1617). Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong.
Prehn, M. (2012). *Vowel quantity and the fortis-lenis distinction in North Low-Saxon*. Utrecht: LOT.
Remijsen, B., Ayoker, O. G., & Mills, T. (2011). Shilluk. *Journal of the International Phonetic Association*, *41*(1), 111–125.
Remijsen, B., & Gilley, L. (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics*, *36*(2), 318–344.
Remmel, M. (1975). *The phonetic scope of Estonian: Some specifications*. Preprint KKI-5. Academy of Sciences of the Estonian S.S.R. Academy of Sciences of the Estonian S.S.R. Tallinn: Institute of Language and Literature.
Ternes, E. (1973). *The phonemic analysis of Scottish Gaelic: Based on the dialect of Applecross, Ross-shire*. Hamburg: Helmut Buske Verlag.
Thomas, K. D. (1992). Vowel length in Yavapai revisited. In: J. E. Redden (Ed.), *Occasional papers on linguistics 17. Papers from the 1992 Hokan-Penutian languages conference and the J. P. Harrington conference held at the Univerity of California, Santa Barbara and the Museum of Natural History, Santa Barbara, June 24–27, 1992* (pp. 90–131).
Thomas, K. D., & Shaterian, A. (1990). Vowel length and pitch in Yavapai. In: J. E. Redden (Ed.), *Occasional papers on linguistics 15. Papers from the 1990 Hokan-Penutian languages workshop held at Univerity of California, San Diego, June 22–23, 1990* (pp. 144–153).
Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, *88*(1), 97–100.
Traunmüller, H., & Krull, D. (2003). The effect of local speaking rate on the perception of quantity in Estonian. *Phonetica*, *60*(3), 187–207.
Vainio, M., Järvikivi, J., & Aalto, D. (2010). Phonetic tone signals phonological quantity and word structure. *Journal of the Acoustical Society of America*, *128*(3), 1313–1321.
van Haitsma, J. D., & van Haitsma, W. (1976). *A hierarchical sketch of mixe as spoken in San José El Paraíso*. SIL Publications in Linguistics and Related Fields, 44. Norman: SIL of the University of Oklahoma.
Viitso, T.-R. (2003). Phonology, morphology and word formation. In: M. Erelt (Ed.), *Estonian language. Linguistica uralica, Supplementary Series*, Vol. 1 (pp. 9–92). Tallinn: Estonian Academy Publishers.
Yu, A. C. L. (2010). Tonal effects on perceived vowel duration. In: C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Laboratory phonology*, Vol. 10 (pp. 151–168). Berlin, New York: Mouton de Gruyter.