

Domain of the Estonian Quantity Degrees: Evidence from Words Containing Diphthongs

Arvo Eek and Einar Meister

Laboratory of Phonetics and Speech Technology
Institute of Cybernetics at Tallinn Technical University
Akadeemia tee 21, Tallinn, EE 12618, Estonia

Abstract

The Estonian quantity degrees (Q) are phonological two-syllable prosodic units the distinct durational patterns of which are based on various combinations of duration ratios of foot-internal neighbouring phonemes. In order to test this definition, the durations of word types $C_1V_1V_2C_2V_3$ (Q2) and $C_1V_1V_2:C_2V_3$ (Q3) separated from a text read as a dialogue with three different speech rates were measured. The discrimination of the duration ratios of successive phonemes and their relevance for the identification of the quantity degrees were examined by perception tests. Listeners cannot determine by merely the durational relationships of the first syllable whether the syllable belongs either to one or the other word form (quantity degree), although they can discriminate whether the coda vowel is longer or shorter than the nucleus vowel. The duration ratios between the stressed syllable coda and nucleus, and the coda and the unstressed syllable nucleus are complementary cues for the perception of Qs.

Introduction

The Estonian quantity degrees are phonological disyllabic prosodic units the distinct durational patterns of which are based on various combinations of duration ratios of foot-internal neighbouring phonemes. Pursuant to this definition the domain of the quantity degree (quantity 1, quantity 2 and quantity 3, i.e. Q1, Q2 and Q3 respectively) is a sequence of a stressed and unstressed syllable of a foot. Various words and morphological forms of the same word (e.g. /kalas/ 'kala, iness.sg., Q1' – /kallas/ 'kallas, nom.sg., Q2' – /kal:las/ 'kallama, imperf. 3.p.sg., Q3'; /kaunis/ 'kaunis, nom.sg., Q2' – /kau:nis/ 'kaunis, iness.sg., Q3') are differentiated on the basis of their degree of quantity. Monosyllabic words do not have the opposition of Q2 – Q3.

Theoreticians who recognize quantity degrees only as syllable size prosodic units [1, 2], consider merely the durational (and tonal) differences of a stressed syllable sufficient for the discrimination of quantity degrees. Meanwhile the reciprocal duration of the second syllable vowel of a foot and the rhyme of the first syllable is a predictable accompanying phenomenon dependent on the first syllable and therefore does not have a distinctive weight for the determination of a phonological unit. Measurements of duration have demonstrated [3] that Q's are characterized by certain duration ratios between the

stressed first and unstressed second syllable; these relations remain stable also in spontaneous speech [4]. However, beside the difficulties of defining a syllable boundary it should be emphasized that both the native Estonian listeners and English listeners react in a similar binary way to the ternary opposition of durational relationships of syllables (2:3, 3:2 and 2:1; Q1, Q2, Q3 respectively): they can discriminate whether the second syllable is longer (Q1) or shorter (Q2, Q3) than the first syllable whereas the gradational difference (3:2 and 2:1) remains indiscriminable [5, 6]. There are also other data throwing doubt to the hypothesis that syllabic duration ratios have a decisive perceptual importance differentiating quantity degrees. It is known that "a lengthening of [sa] in the first syllable of [sat-tu] had an effect similar to that of lengthening the [u] of the second syllable: it made [t] seem shorter [7, 8]. According to the syllabic ratio hypothesis, the effect of lengthening [sa] should be opposite to that of lengthening [u]". These results make us really discard the treatment of syllabic duration ratios while the question about the domain of quantity degrees remains still unanswered. Rather, it would be more appropriate to say that the last referred data about the effect of the duration of the second syllable vowel on the determination of Q boundaries support our understanding that the domain of Qs is in its minimum form a sequence of two syllables. The hypothesis according to which Qs are identified by comparing the durations of neighbouring phonemes (not syllables!) in synchrony with the speech flow within a foot of two successive syllables (cf. also [9]) will be examined below.

Results and discussion

We have measured two-syllable word forms containing diphthongs separated from our dialogue corpus Tempo spoken in different speech rates (Table 1). Table 1 demonstrates that in the Q2 and Q3 feet C_1 , V_1 and C_2 do not differ by duration. The differences of the ratios V_2/V_1 and V_3/V_2 strike out as essential, remaining more or less constant for all speech rates and in all examined contexts. Previous measurements of monosyllabic words CV(V)C(C) (dialogue corpus Tempo, 600 words, 4 speakers) revealed that a vowel can be classified as a short one (single vowel) if $V \leq 1.4 \times C_1$ and as a long monophthong (double vowel) if $VV > 1.4 \times C_1$, where $C_1 = /p/$ with average durations ca 70, 100 and 130 ms in fast, moderate and slow rate respectively.

Table 1. Average durations and duration ratios of phonemes in /lautə/ ‘laut, gen.sg., Q2’ and /lau:ta/ ‘laud, part.sg., Q3’ separated from dialogue corpus spoken in fast (F), moderate (M) and slow (S) speech rate (32 words per rate, 96 words per Q, total — 192 words, 4 speakers).

Q	Tempo	C ₁	V ₁	V ₂	C ₂	V ₃	V ₂ /V ₁	V ₃ /V ₂	
Q2	F	\bar{x}	43	68	53	45	95	0.8	1.8
		σ	11.8	9.1	8.5	9.5	16.7		
	M	\bar{x}	58	96	76	60	120	0.8	1.6
		σ	18.6	16.1	11.6	9.2	13.8		
	S	\bar{x}	91	160	117	93	194	0.7	1.7
		σ	24.9	20.8	23.2	13.7	26.7		
Q3	F	\bar{x}	39	73	100	48	72	1.4	0.7
		σ	10.0	13.0	16.6	11.0	13.6		
	M	\bar{x}	56	108	143	62	87	1.4	0.6
		σ	12.7	19.8	21.2	7.2	12.2		
	S	\bar{x}	79	160	236	85	126	1.5	0.5
		σ	13.2	26.1	40.9	11.4	23.7		

Using this type of monosyllabic stimuli (where the vowel was altered within a time interval of 40-280 ms by three speech rate imitations of C₁ 70, 100, 130 ms; 90 stimuli), 20 listeners were asked to identify in the perception test whether the word they heard was *sas* or *saas*. The longer the onset consonant C₁ was, the longer was the vowel duration marking V-VV boundary. The duration ratio of V/C₁ remained approximately in concordance with the classification data of the read text, yet differences appeared for slow speech rate where the long absolute duration of the vowel may have affected the determination of the boundary so that the V/C₁ ratio for the double vowel appeared to be smaller than stated above (VV ≥ 1.3xC₁). However, the big differences of intrinsic durations of

consonants create obstacles for the manipulation of duration ratios. The average durations of the onset consonant C₁ = /l/ in Table 1 are in three speech rates ca 40, 60 and 80ms, these are 30, 40 and 50ms shorter than C₁ /p/ with the respective speech rates. Thus, all first components of the diphthong (V₁) in Table 1 would be classified as long vowels on the basis of V₁/C₁ ratio. We assume that at the period of language acquisition the differences of automatic intrinsic durations are learned to be taken into account. At the process of normalisation (i.e., in the present case, adding the mentioned differences to the duration of /l/) all first components of the diphthong will be classified in all speech rates as short vowels (the largest V₁/C₁ ratio is 1.2) and all V₁+V₂ cases would belong to the same group with long monophthongs (the smallest V₁+V₂/C₁ ratio is 1.7). The ratios V₂/V₁ and V₃/V₂ have an extremely strong classification capacity for keeping Q2 and Q3 feet separate (see the plotting of single cases in Fig. 1). In order to confirm the syllabic nature of quantity degrees the auditory distinction of the oppositions of measured durational differences of the components of diphthongs should be proved. According to the results by D.Klatt [10] only a 20-25% or even larger duration change can be discriminated.

The discrimination threshold of two neighbouring phonemes has been set in the present report on the difference of 20%. With such a criterion the durations of V₁ and V₂ of Q2 words remain probably in 58% cases indistinguishable. However, the cases which remain probably indistinct on the basis of V₂/V₁ reveal a sufficiently strong difference between V₃ and V₂ so that finally the durations of V₂ and V₁, also V₃ and V₂ will be indistinguishable only in 4% of all cases.

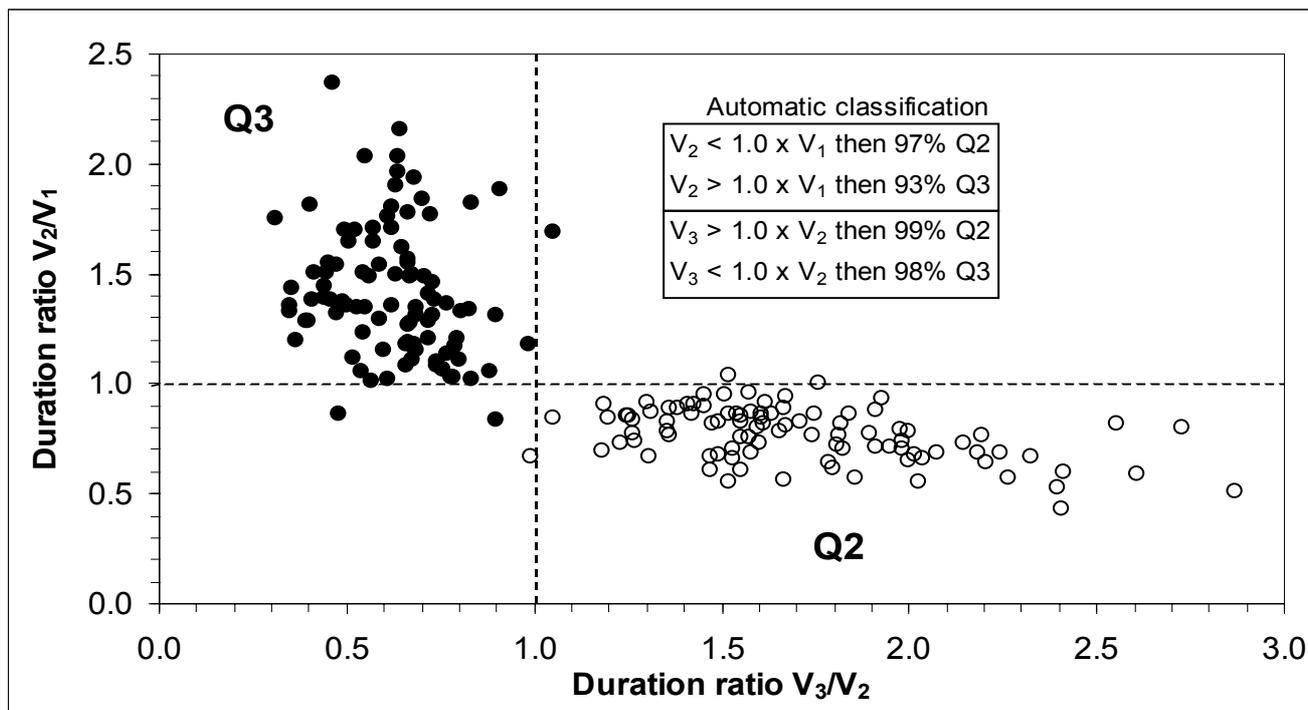


Figure 1. The words /lautə/, Q2 and /lau:ta/, Q3, read in fast, moderate and slow speech rate. Testing differences of quantity degrees on the basis of V₂/V₁ and V₃/V₂ (192 words, 4 speakers).

V_1 and V_2 cannot be differentiated in Q3 words probably for 31% of cases, yet the majority of these are compensated with the perceptibly large V_3/V_2 ratio so that both pairs of vowels will remain indistinguishable by duration in 12% of the words. There is a tendency in Q2 words /lauta/ to keep the first syllable rhyme vowels relatively equal (i.e. indistinguishable by duration, preventing the prominence of coda) and the nucleus vowel of the second syllable not shorter than the coda of the first syllable (the so-called balance, dispersion). The Q3 words /lau:ta/ have a tendency to keep V_2 longer than V_1 and the nucleus of the second syllable shorter than the coda V_2 of the first syllable (the so-called imbalance, centralisation). A perception test 1 was arranged for the examination of the relevance of both syllables of the foot in the identification of Qs (144 stimuli, 35 listeners, stimulus word *suigu* where C_1 , V_1 and C_2 were unaltered with the respective durations 90, 100 and 70 ms; V_2 was altered within 40-210 ms for four durations of V_3 (50, 80, 130, 180 ms); F_0 monotonous 130 Hz; listeners were asked to decide whether the word they hear, *suigu*, is a place name, nom.sg., Q2 or a place name, illat.sg., Q3). The experiment confirms the relevance of both syllables for the identification of Qs. If the duration of V_2 was sufficiently longer than that of V_1 (at least 1.4-1.5 times), Q3 was identified even for the indistinguishable V_3/V_2 ratio (i.e. when V_3 was relatively long); for a shorter V_3 , Q3 was identified also in these cases when V_1 and V_2 were equal by duration. If the duration of V_3 is very long, the Q2-Q3 boundary is determined by V_2/V_1 alone: if it is ≥ 1.4 , we have Q3, if < 1.3 , then we perceive Q2. The stimuli which are identified as Q2 do not include V_2 that should be perceived longer than V_1 ; short V_3 decreases the amount of Q2 decisions.

Let us have an extract from test 2, which proves that Q2 – Q3 are identifiable also in this case when the rhymes of stressed first syllable are fully identical. F_0 peak that marks the stress was shifted from the first syllable to the second syllable in the word *kanata* /kanatta/ ‘kana, abess.sg., Q1’ (where the durations of C_1 , V_1 , C_2 , V_2 , C_3 , V_3 are respectively 80, 65, 60, 85, 200, 50 ms, F_0 peak 147 Hz on the first syllable), this gave the result that the word /kanatta/ was perceived as an unstressed auxiliary word /ka/ + /nat:ta/ ‘natt, part.sg., Q3’; when V_3 was lengthened up to 140-150 ms, the stimulus was perceived as an auxiliary word /ka/ + /natta/ ‘natt gen.sg., Q2’. The shorter V_3 creates an auditory illusion of a longer coda of the previous syllable (Q3) and the longer V_3 creates an illusion of a shorter coda (Q2). Such decisions of the listeners depend merely on perceptibly different duration ratios between V_3 and C_3 (their decisions depend also on the location of comparable neighbouring segments regarding the stress peak). The duration ratios between the coda and nucleus of a stressed syllable, and the coda and nucleus of an unstressed syllable are complementary cues for the perception of quantity degrees.

The results of perception test 3 are presented in Figure 2; the second syllable *gu* was removed from the test 1 words *Suigu* and the listeners were asked to decide whether the remaining first syllable *sui-* belongs to the word *suigu* in

nominative case (Q2) or in illative case (Q3). The results were chaotic demonstrating that the listeners could not determine the origin of the syllable from either one or the other word form (quantity degree) merely on the basis of the duration ratios within the first syllable.

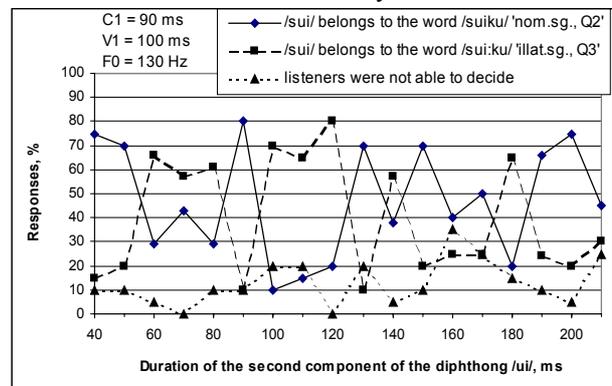


Figure 2. Recognition of the words *Suigu* ‘nom.sg., Q2’ and *Suigu*, illat.sg., Q3’ by duration ratios of the components V_2/V_1 of the first syllable diphthong (21 listeners).

In test 4 the listeners were asked to estimate whether the second component V_2 of the diphthong in the syllable *sui* is shorter or longer than the first component V_1 (Fig. 3). V_2 was perceived shorter by the majority of listeners when $V_2/V_1 \leq 0.8$; there were also slightly more responses which stated V_2 to be shorter when both components should have been indistinguishable. V_2 was perceived longer than V_1 when $V_2/V_1 \geq 1.4$. Despite the fact that listeners discriminate the prototypical Q2 – Q3 intrasyllabic relations $V_2 < V_1$ and $V_2 > V_1$, they still cannot identify quantity degrees by the duration ratios of one syllable.

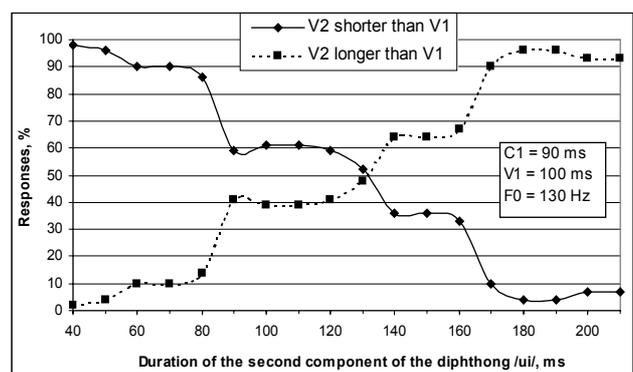


Figure 3. Perception of duration ratios V_2/V_1 of the diphthong in the syllable *sui* (the same stimuli as in test 3; increasing and decreasing order of $V_2 - 18 \times 2 = 36$ stimuli, 21 listeners).

Three kinds of syllables should be distinguished in the Estonian language for proper generation of word stress patterns. In the metrical stress theory [11], which *mutatis mutandis* suits for the description of Estonian stress system, the syllables defined by their weight are called light, heavy and superheavy. A light and a heavy syllable can occur in a metrically strong and weak position of the foot, a superheavy syllable only in a strong position of the

foot. A light syllable is monomoraic and a heavy and superheavy are bimoraic syllables. The distinction between a heavy and superheavy syllable is based on the mora sharing between the nucleus and coda in the way presented in this report. We consider it partly correct that light and heavy syllables are both monomoraic and only a superheavy syllable is bimoraic [2, 12]. Such treatment is based on the phenomenon that a stressed light or heavy syllable has to be followed by at least one unstressed syllable in order to form a foot whereas a superheavy syllable may be followed by a new stressed syllable. In stress metrics it is not possible to consider a heavy syllable equal to a light syllable as then there is not possible to describe a restriction caused by a heavy syllable in the situation where in the footing process the primary strong local parsing can freely vary with weak local parsing: if the syllable σ_{n+2} is heavy, the stress assigning rule does not allow to skip the syllable and only the strong local parsing appears to be possible – e.g. (*vallut*)(*tattut*)(*tet:ta*) ‘*vallutatu*, abessive pl.’; however, in case of a light syllable both modes of footing are possible – e.g. either (*vallut*)(*tajat*)(*tet:ta*) ‘*vallutaja*, abessive pl.’ or (*vallut*)*ta*(*jattet*)*ta*.

A syllable does not have a quantity degree. It has never had it. Long syllables became differentiated as a result of foot isochrony: the coda of the first syllable became shorter (heavy syllable) before the closed second syllable, and before the open second syllable the coda either became longer or seemed longer (superheavy syllable) before the extra-short second-syllable vowel. Strengthening of the stress (perhaps due to the influence of Germanic contacts?) turned the superheavy syllable into a monosyllabic word after the apocope (and syncope) of the extra-short vowel of the open second syllable. Such monosyllabic words have never had been in Q3. Heavy and superheavy syllables originated earlier than quantity degrees, therefore it is not correct to speak about the Q2 – Q3 neutralisation in monosyllabic words nor about a Q3 archidegree. It is not necessary to define in a morpho-phonological and morphological analysis a monosyllabic word as being in Q3. The opposition heavy – superheavy is sufficient: as a monosyllabic word is superheavy in the nominative case, the first syllable will be superheavy also in the partitive case or if the first syllable is heavy in the nominative case, it will be also heavy in the partitive case (e.g. /lau:l/ ‘*laul*, nom.sg., superheavy monosyllabic word’ and /lau:lu/ ‘*laul*, part.sg., superheavy syllable + light syllable = Q3’; /mõtte/ ‘*mõte*, nom.sg., heavy first syllable’ and /mõttett/ ‘*mõte*, part.sg., heavy syllable + heavy syllable = Q2’). The opposition Q2 – Q3 supposedly arose after the loss of a word-final consonant and the shortening of the contracted vowel [12] (e.g. [saūnan] > [saūna] > /sauna/ ‘*saun*, gen.sg., Q2’ and [saūnäta] > [saūnä] > /sau:na/ ‘*saun*, part.sg., Q3’); this brought along massive arise of homonyms of morphological forms and under their pressure the above mentioned differences in durational patterns, based on foot isochrony, could have become stronger and phonologize into quantity degrees which in the present day language are disyllabic

suprasegmental units distinguishing words and word forms.

A light, heavy and superheavy syllable are basic elements of the stress structure of a word. However, a superheavy and heavy syllable (see Fig. 1) need not be always differentiated. It does not create problems for monosyllabic words as a new word can always follow. A speaker has two relatively free options for the pronunciation of multisyllabic words in Estonian: a superheavy syllable may be directly followed by stress or not. In the first case a superheavy syllable is treated as a disyllable in metrical stress theory [11] and its pronunciation is more marked (e.g. (*kau:*)(*kele*)*ki* ‘*kaugelegi*’) than in the case when a superheavy syllable is treated as a single syllable forming a disyllabic trochee with the following syllable (e.g. (*kau:ke*)(*leki*)); in the latter occasion (i.e. in Q3 foot) the duration ratio between the nucleus of the unstressed syllable and the coda of the previous syllable can make a superheavy syllable more distinct (as we know from perception tests 1 and 2).

References

- [1] M.Hint. “Why syllabic quantity? Why not the foot?” Proceedings of the Finnic Phonetics Symposium. *Linguistica Uralica* 1998, 34, 3: 172-177.
- [2] M.Ehala. “Eesti väldete probleemi üks lahendus.” *Keel ja Kirjandus* 1999, 6: 378-385, 7: 453-465.
- [3] I.Lehiste. “Segmental and syllabic quantity in Estonian. American Studies in Uralic Linguistics. Indiana University Publications, Uralic and Altaic Series 1: 21-82. Bloomington.
- [4] O.Engstrand, D.Krull. “Durational correlates of quantity in Swedish, Finnish and Estonian: cross language evidence for a theory of adaptive dispersion.” *Phonetica* 1994, 51: 80-91.
- [5] R.A.Fox, I.Lehiste. “Discrimination of duration ratios by native English and Estonian listeners.” *Journal of Phonetics* 1987, 15: 349-363.
- [6] R.A.Fox, I.Lehiste. “Discrimination of duration ratios in bisyllabic tokens by native English and Estonian listeners.” *Journal of Phonetics* 1989, 17: 167-174.
- [7] D.Krull, H.Traunmüller. “Perception of quantity in Estonian.” Proceedings of Fonetik 2000. University of Skövde, Department of Languages. 2000: 85-88.
- [8] D.Krull, H.Traunmüller. “Perception of quantity in Estonian (Part II).” Royal Institute of Technology, Speech Transmission Laboratory, Quarterly Progress and Status Report 43, 2002.
- [9] W.A.van Dommelen. “Auditory accounts of temporal factors in the perception of Norwegian disyllables and speech analogs.” *Journal of Phonetics* 1999, 27: 107-123.
- [10] D.H.Klatt. “Linguistic uses of segmental duration in English: acoustic and perceptual evidence.” *Journal of the Acoustical Society of America* 1976, 59: 1208-1221.
- [11] B.Hayes. *Metrical stress theory: principles and case studies*. University of Chicago Press. Chicago 1995.
- [12] M.Hint. *Eesti keele astmevahelduse ja prosoodiatüübi tüpoloogilised probleemid*. Tallinn, Helsinki 1997.