# SIGNIFICANT FEATURES IN ESTONIAN WORD PROSODY

*Meelis Mihkla & Mari-Liis Kalvik*

Department of Language Technology, Institute of the Estonian Language, Tallinn, Estonia
meelis@eki.ee; mariliis@eki.ee

## ABSTRACT

The article discusses the main paradigm of Estonian word prosody – triple opposition of phonetic quantity in disyllabic foot – as manifested in fluent speech. Statistical methods are used to find out which of the characteristics of quantity degrees are essential and on what conditions they occur. The word structure under investigation is CV(::)CV.

As revealed by statistical analysis, the duration ratio of the stressed vs. unstressed syllable is still the primary feature distinguishing between the Estonian quantity degrees. The tonal component, pitch, is also important, but instead of the characteristics of the tonal peak (or turning point TP), which is changeable and malleable by other factors, the most significant of the possible alternative features should rather be seen in the ratio of the mean F0 values of the stressed and unstressed syllables. Another sufficiently relevant feature is the position of the word in the phrase (final vs. non-final), while different models apply to content and function words. Intensity and stressedness of words turned out to be marginal for the given material.

**Keywords:** quantity degree, durational and tonal characteristics, Estonian word prosody

## 1. INTRODUCTION

Quantity degrees – short Q1, long Q2 and overlong Q3 – belong to the key entities of the Estonian phonetic system. In essence these are suprasegmentals realized in the primary stressed disyllabic foot. Quantities are differentiated by both temporal (primary) and tonal (secondary) characteristics and they have a phonemic function. The temporal parameter is represented by the duration ratio of the word's stressed and unstressed syllable:

$$durational\ ratio = \frac{\sigma_{stressed}(nucleus+[coda])}{\sigma_{unstressed}(nucleus)} \quad (1)$$

This feature has hitherto been the most stable feature to distinguish between the Estonian quantity degrees. Numerically, various experiments carried out during the past fifty years have established the following ratios: 2:3 for the short degree (Q1), 3:2 for the long degree (Q2) and 2:1 for the overlong degree (Q3), see, e.g., the fundamental research [13]. In our researches [10, 11] those durational relations between the stressed and unstressed syllables of a word covered three-fourths of the variability of the data. Consequently, any other parameters can have but a specifying role, manifested only in certain lexical paradigms or in special cases. When applying statistical methods to large corpora it is sometimes possible to detect some small, hidden, but still relevant effects of inputs on output [16]. In the present study an attempt is made to determine the relevance of different characteristics and to model the triple opposition of the Estonian quantity degrees by means of linear regression and the CART technique. The study is a summarizing follow-up of a series of earlier articles.

Several earlier studies [6, 13, 14] as well as perception tests [5, 7, 15] have established that the position of the tonal peak in the stressed syllable is a crucial feature to distinguish between the two long quantities: Q2 and Q3. In Q1 and Q2 words the peak is near the syllable boundary of the stressed vowel, whereas in Q3 words it falls on the first third of it. According to our results, in fluent speech the TP of the stressed syllable is often hard to detect (in the rhyme of the stressed syllable the pitch is either smoothly rising or falling, if not completely flat, or it may even have shifted on to the unstressed syllable). Therefore the pitch parameter tested in this study is the ratio of the mean pitch values of the vowels of stressed and unstressed syllables. In addition, there are studies [7] where a higher intensity of the speech signal has been pointed out as a possible marker of Q3. So, like in the pitch feature just described, we have also chosen to test the ratio of the intensities of the stressed and unstressed syllables.

In addition, such characteristics as position of the word in the prosodic phrase, number of syllables in the word, and stressedness of the word are suggested as potential ancillary variables. Accentuation is a condition which is supposed to contribute to the quantity-specific pitch movement in the word. F0 curves in general occur in positions affording more time for their realization [9] – stressed units have greater energy, they tend to be louder and longer, although pitch changes can still occur independently of stress changes [12].

## 2. MATERIAL AND METHOD

The research material consists of 736 words from a corpus of fluent speech (300 Q1, 236 Q2 and 200 Q3 words), consisting texts read out in paragraphs by 13 men and 14 women. Part of the material comes from the Babel corpus [8], part from longer radio news read by professional newsreaders.

Only those words were selected for examination where both the syllable carrying the main stress and the syllable immediately following it had the CV(::)CV structure. Considering Estonian word structure, the duration ratio of the two syllables is traditionally described as a vowel ratio (V1:V2). In Q1 words V1 is short (e.g. *pole* [pole] 'is, are not'), while in Q2 and Q3 words it is either long (e.g. *poole* [po:le] 'half GenSg') or overlong *poole* [po::le] 'towards'). Most of the words in the material have at least two syllables. Words (both content and function words) occupy different positions in the prosodic phrase, some being stressed, some unstressed.

Phonetic analysis was done using the Praat program [4]. Measurements for each word included segment lengths (ms), fundamental frequency (F0) values at the initial ($F0_i$) and final ($F0_f$) boundaries of V1 and V2, and ($F0_{tp}$) at the turning point $t_{tp}$ (TP), where F0 displays a noticeable fall. In addition we measured the distance of the TP from the onset of V1 ($t_{tp} - t_{V1i}$), yielding the rise of F0. The F0 rise % was found as follows:

$$F0_{rise\%} = 100 * \frac{(t_{tp} - t_{V1i})}{(t_{V1f} - t_{V1i})} \tag{2}$$

The ratio of the mean F0 values of the stressed and unstressed syllables was calculated by the formula:

$$F0_{V1/V2} = \frac{2 * (F0_{V1i} + F0_{tp} + F0_{V1f})}{3 * (F0_{V2i} + F0_{V2f})} \tag{3}$$

## 3. RESULTS AND DISCUSSION

The averaged results of the phonetic analysis were as follows: The duration ratios (V1:V2) were 0.8 for Q1, 1.9 for Q2 and 2.8 for Q3. The $F0_{rise\%}$ value was 100 for Q1, 71 for Q2 and 48 for Q3. Both the duration ratio and the $F0_{rise\%}$ divide the words quantity-wise into three distinctive groups according to the analysis of variance (ANOVA) ($p<.0005$ in all cases). The first vowel of a Q1 word has a pitch contour where the TP lies on the syllable boundary, while in Q2 words it is found in the right half of the first vowel and in Q3 words its position is around the middle of the first vowel. The $F0_{rise\%}$ values (indicating the position of the crucial TP) of Q2 and Q3 words do not show a big difference but according to variance analysis it is still significant. The supporting feature is a steep F0 fall from the TP to the syllable boundary, which is characteristic of Q3. In our data the fall for Q3 is 3.1 st. The other readings are 1.2 st for Q2 and 0.8 st for Q1.

The results of statistical analysis are presented in Table 1 above, which demonstrates the selected parameters and their significance as quantity features for the whole material as well as depending on the position of the word in the prosodic phrase (final vs. non-final) and on the function of the word (content vs. function word).

**Table 1:** Factors affecting the quantity degree and their significance for the whole material, for female and male speakers as well as depending on the position of the word in the prosodic phrase and on the function of the word.

| | No. of words | Squared multiple R | V1/V2 | $F0_{rise\%}$ | Position of word in phrase | Accent | Intensity ratio | $F0_{V1/V2}$ |
|---|---|---|---|---|---|---|---|---|
| Total | 607 | 0.772 | + | + | – | – | – | + |
| Female speakers | 262 | 0.745 | + | – | – | +/– | – | + |
| Male speakers | 345 | 0.793 | + | + | – | – | – | +/– |
| Phrase-final words | 92 | 0.857 | + | +/– | – | – | +/– | – |
| Non-phrase-final words | 515 | 0.766 | + | + | – | + | – | + |
| Content words | 518 | 0.760 | + | + | + | – | – | + |
| Function words | 88 | 0.822 | + | – | – | – | – | – |

The results of statistical analysis are presented in Table 1 above, which demonstrates the selected parameters and their significance as quantity features for the whole material as well as depending on the position of the word in the prosodic phrase (final vs. non-final) and on the function of the word (content vs. function word). Also, separate analyses have been done on the words pronounced by male and female speakers.

Significance of the features was tested by means of regression analysis. The significant argument features have been marked by a + sign (in all those cases p<.05), while the features insignificant in the given situation (p>.1) bear a minus sign, while +/– designate the conventional 'grey zone' between significance and insignificance (.05<p<.1).

The results of the analysis are as follows. The duration ratio (V1:V2) as the classical quantity feature expectedly shows a stable positive occurrence in all situations examined. The significance of the pitch-based parameters $F0_{rise\%}$ and $F0_{V1/V2}$ is not so absolute, but mutually more or less equal, with slight variation in special cases. The result gives reason to test the ratio of mean F0 values of the foot as an alternative feature in future studies. The exact position of the word in the phrase is significant only in the content word model. In the rest of the cases the binary variable (phrase-final or not) seems to suffice.

The accentuation of the word is significant only in modelling the quantity of non-phrase-final words. The result may be related to the fact that in our material accented words are usually not located at the end of the phrase. In Estonian declarative sentences the usual intonation contour is H*+L [3]. Our data consists of news or news-like sentences and they are in general read as messages where the theme can but does not need to be marked with accentuation. Individual accentuations of course exist but their amount is not remarkable. There are also utterances with the so-called *totally downstepped* last stressed syllable in the phrase – such words have a flat and low F0 contour [2]. In such cases the tonal feature is not at work and the quantity distinguishing between Q2 and Q3 is based on the duration ratio only. There is also another point: as the determination of stressedness is by experience subjective and variable, the method of multiple experts should be used in future studies.

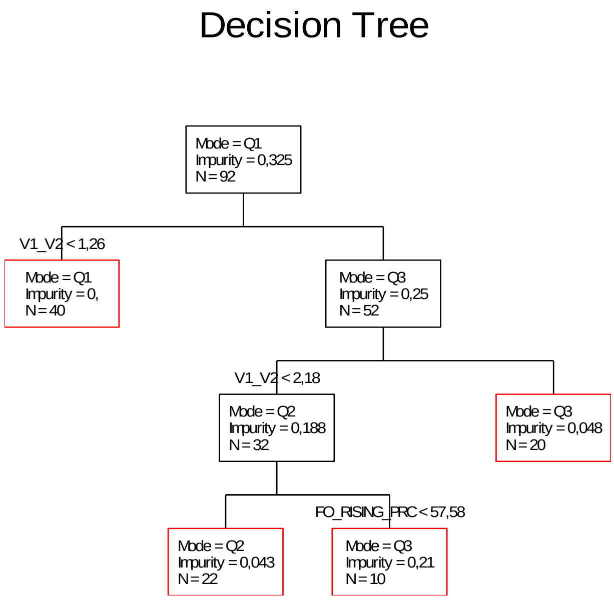
The intensity ratio, however, turned out to be marginal in most cases, having a certain influence only on the quantity opposition of phrase-final words. This may serve as a hint that intensity and F0 can act independently, at least to some degree, although duration, intensity and F0 are not supposed to be autonomous from one another neither in the process nor in the perception of speech [1].

Although there are also certain differences in the male vs. female results (possibly due to their quite different F0 values), it would be premature to declare an immanent gender difference, at least not before testing more readers from both genders.

The simplest model covered the group of function words, using duration ratio as the only significant feature. The reason may lie in the simpler structure of function words and in the higher number of uninflected words in this group.

The rest of the variable candidates tested (phrase length, compounding, duration of syllables and parts of syllables etc.) turned out to be invariably insignificant.

**Figure 1:** Regression tree for determining the phonetic quantity (Q1, Q2, Q3) of the final word of a prosodic phrase.




The above regression tree (Fig. 1) shows the description of quantity opposition in phrase-final words. In a general case phrase-final words are pronounced longer than those in any other position (the final lengthening phenomenon), while their pitch can be low and flat. Nevertheless, their final position does not change the durational relations determining their quantity. The duration ratio functioning as the threshold between the words of the first quantity degree (Q1) and the rest is 1.26.

If, however, the V1:V2 ratio is at least 2.18, we have, no doubt, a Q3 word. The intermediate values need to be sorted further by another argument feature, which is the $F0_{rise\%}$: If the rise lasts less than 57.58% of the duration of V1, we have a Q3 word, in the opposite case the quantity is Q2.

## 4. CONCLUSIONS

Our analysis of Estonian word prosody in fluent speech has led to the following conclusions: The most important parameter enabling one to detect the phonetic quantity of words pronounced in fluent speech and to model quantity oppositions is the duration ratio of the stressed and unstressed syllable. This feature (expressed here as the ratio of the vowels of stressed and unstressed syllables V1:V2) appears significant in all experiments. Pitch (F0) contour is also a significant feature but the mechanism (combination of conditions) of its occurrence and contribution is not clear. As the tonal peak need not always be detectable in the stressed syllable one can use the ratio of the mean F0 values of the stressed and unstressed syllables as a practically equivalent feature. The parameter needs to be tested and further specified using additional material. Another relatively important feature is the position of the word in the phrase (final vs. non-final), which also needs future research. Also, content and function words turned out to require different quantity models. This suggests that it may be necessary to take a closer look, as far as quantity is concerned, at different parts of speech. The intensity and accentuation of a word were not proved significant features for the material analysed in the context of this study.

## 5. ACKNOWLEDGEMENTS

This work has been supported from the grant ETF7998 and the project SF0050023s09.

## 6. REFERENCES

[1] Andreeva, B., Barry, W., Steiner, I. 2007. Producing phrasal prominence in German. *Proc. 16th ICPhS* Saarbrücken, 1209-1212.
[2] Asu, E.L. 2001. An autosegmental-metrical analysis of Estonian intonation: preliminary results. In Seilenthal, T. (ed.), *Proc. of the IXth ICFU* Tartu, 83-89.
[3] Asu, E.L., Nolan, F. 2001. The interaction of intonation and quantity in Estonian: an analysis of nuclear falls in statements and questions. *Proc. of the VIIIth Conference Trondheim 2000* Frankfurt, 23-32.
[4] Boersma, P., Weenink, D. 2010. Praat: Doing phonetics by computer [computer program], *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. 1983. Kvantiteet ja rõhk eesti keeles (I). Fonoloogiliste tõlgenduste kriitikat. *Keel ja Kirjandus* 9. 481-489.
[7] Eek, A., Meister, E. 1997. Simple perception experiments on estonian word prosody: Foot structure vs. segmental quantity. In Lehiste, I., Ross, J. (eds.), *Estonian Prosody: Papers from a Symposium.* Tallinn: Institute of Estonian Language, 77-99.
[8] Eek, A., Meister, E. 1999. Estonian speech in the BABEL multi-language database: Phonetic-phonological problems revealed in the text corpus. *Proc. of LP 1998*, 529-546.
[9] Gussenhoven, C. 2000. The lexical tone contrast of roermond Dutch in optimality theory. In Horne, M. (ed.), *Prosody: Theory and Experiment.* Dordrecht, Boston, London: Kluwer Academic Publishers, 129-167.
[10] Kalvik, M-L., Mihkla, M. 2010. Modelling the temporal structure of Estonian speech. In Skadina, I., Vasiljevs, A. (eds.), *Human Language Technologies. The Baltic Perspective: Proc. 4th IC.* Amsterdam, 53-60.
[11] Kalvik, M-L., Mihkla, M., Kiissel, I., Hein, I. 2010. Estonian: Some findings for modelling speech rhythmicity and perception of speech rate. In Sojka, P., Horak, A., Kopecek, I., Pala, K. (eds.), *Text, Speech and Dialogue*. Berlin/Heidelberg: Springer Verlag, 314-321.
[12] Ladefoged, P., Johnson, K. 2010. *A Course in Phonetics* (6th ed.). USA: Wadsworth, Gengage Learning.
[13] Lehiste, I. 1960. Segmental and syllabic quantity in Estonian. *American Studies in Uralic Linguistics* Bloomington: Indiana University, 1, 21-28.
[14] Liiv, G. 1961. Eesti keele kolme vältusastme kestus ja meloodiatüübid. *Keel ja Kirjandus* 7-8, 412-424; 480-490.
[15] Lippus, P., Pajusalu, K., Allik, J. 2011. The role of pitch cue in the perception of the Estonian long quantity. In Frota, S., Elordieta, G., Prieto P. (eds.), *Prosodic Categories: Production, Perception and Comprehension*. Springer, 231-242.
[16] Sagisaka, Y. 2003. Modeling and perception of temporal characteristics in speech. *Proc. 15th ICPhS* Barcelona, 1-6.