An acoustic study of Estonian word stress

Pärtel Lippus [1,2], *Eva Liina Asu* [1], *Mari-Liis Kalvik* [1,3]

[1] Institute of Estonian and General Linguistics, University of Tartu, Estonia
[2] Institute of Behavioural Sciences, University of Helsinki, Finland
[3] Institute of the Estonian Language, Estonia

partel.lippus@ut.ee, eva-liina.asu@ut.ee, Mari-Liis.Kalvik@eki.ee

Abstract

This study investigates the acoustic correlates of word stress in Estonian. It forms part of a broader international collaboration the aim of which is to develop a universal language-independent model for evaluating lexical stress regardless of the phonological structure of a given language. To this aim the characteristics of word stress in a range of languages are studied using unified methodology. For the present study, four acoustic measures were analysed as a function of speaking style and stress: vowel duration, F0 mean, F0 standard deviation, and spectral emphasis. The results show that the strongest correlate of style and stress in Estonian is vowel duration, but stress has a strong interaction with the Estonian three-way quantity system.

Index Terms: word stress, quantity, speaking style, Estonian

1. Introduction

The present study forms part of the Word Stress Project (http://wordstress.ling.gu.se/index2.html), which addresses the acoustic description of word stress correlates and the perception of these correlates as cues to relative syllable prominence. At present 7 languages are included: Brazilian Portuguese, British English, Estonian, French, German, Italian and Swedish. The results of the project will ultimately be used to formulate a typology of word stress suggesting methods that could be applied to the study of any language.

All languages included in the project are studied using the same methodology, whereby the acoustic correlates of stress are described as a function of stress level (primary, secondary and unstressed) and speaking style (spontaneous, phrase reading and word list reading). The acoustic correlates measured for each syllable of a word are the mean F0, standard deviation of the F0, vowel duration, and spectral emphasis. So far, results have been obtained for Swedish [1], [2], Brazilian Portuguese [3], [4], and German [5]. In this paper similar methodology will be applied to Estonian – an unrelated Finno-Ugric language which is known for its three quantity degrees.

In Estonian, the main word stress is usually fixed on the first syllable. It can be elsewhere only in foreign and loan words, interjections and names. As pointed out by Lehiste, the role of stress in Estonian is primarily identificational rather than contrastive [6]. The position of stresses, and thus the division of the word into feet, follows the trochaic rhythm structure. Disyllabic or trisyllabic words typically consist of a single foot (a primary stressed syllable is followed by one or two unstressed syllables). A tetrasyllabic word is generally made up of two disyllabic metric feet. Secondary stresses normally fall on successive odd-numbered syllables (e.g. *magamata* [ˈmɑ.kɑ.ˌmɑtː.tɑ] 'sleepless', *lõpetatigi* [ˈlɤp.pe.ˌtɑt.ti.ki] 'was finished'). In longer words the stress pattern is not fully predictable, although it is often determined by the morphology.

All primary stressed syllables in Estonian are in one of the three quantity degrees: short (Q1), long (Q2) and overlong (Q3). This distinction operates only over a disyllabic trochaic foot, and the decisive factor in determining the degree of quantity is the duration ratio between the first (stressed) and second (unstressed) syllable, regardless of the stressed syllable structure [7]. An additional cue to the quantity distinction is the pitch contour which in Q1 and Q2 is a step down between the two first syllables, but in Q3 is realised as an early fall in the first syllable.

Word stress in Estonian has been subject to thorough phonological investigations [8]–[10] whereas very little is known about its phonetic characteristics. Earlier studies addressing acoustic correlates of Estonian stress are limited to a small-scale investigation into the influence of stress on nasal flow, amplitude and duration [11], where duration was shown to be the parameter on which stress had most consistent influence. The temporal patterns of stressed and unstressed syllables have also been studied in relation to the quantity system. A study of penta- and hexasyllabic words [12] showed that the secondary stressed feet were systematically shorter than the stressed feet while the stressed-to-unstressed syllable ratio was retained to a certain extent. Additionally, there is some acoustic evidence which shows that the degree of stress in Estonian has an influence on vowel quality [13], [14].

The aims of the present study are two-fold: firstly, to test the suitability of the methodology of the Word Stress Project on Estonian, which is a typologically different language from the others included in the project so far, and secondly, to gain new knowledge about the little-studied acoustic correlates of word stress in Estonian. As the main stress in Estonian is fixed it could be hypothesised that stress correlates are fewer and not as strong as in languages with a morphologically unbound stress like, for instance, English. We predict vowel duration to be the most important correlate.

2. Materials and method

Identical methodology as used by the other studies carried out in the framework of the international Word Stress Project (e.g. [1], [3], [5]) was applied. The starting point of the materials was the University of Tartu Phonetic Corpus of Spontaneous Speech (http://www.keel.ut.ee/foneetikakorpus). Spontaneous recordings of 16 native speakers of Estonian – 10 females and 6 males – were chosen from the corpus on the basis of the speaker's age and dialectal background. All speakers were between 22 and 34 years old, and represented the Standard North Estonian variety. At the time of the recording they were either students or graduates of the University of Tartu. From each (approximately 30 min long) recording 15–20 utterances were selected per speaker, and from each utterance one or two target words were chosen. The two criteria for selecting the

10.21437/SpeechProsody.2014-34

words were the position of the word in the utterance (prenuclear i.e. not preceded or followed by an IP-boundary) and the number of syllables (at least two or preferably more syllables long). The words therefore display various syllable structures and phonological quantities. Two lists were prepared with these utterances and target words where each phrase and word appeared three times in random order. The two lists were read by the same 16 speakers.

The final data set comprises 276 target words in three different speaking styles: spontaneous speech, phrase reading and word list reading. Most commonly, in our data, a target word is trisyllabic.

The recordings were analysed in Praat [15]. The analysis focussed on the vowels of the target words which were manually labelled. Combined labels were given for stress and quantity. Stress was marked with numbers 3 (primary), 2 (secondary), and 1 (unstressed), and depending on the quantity of each foot the relevant number was used once (Q1), twice (Q2) or three times (Q3), as for instance: 3+1+1 in case of a Q1 foot (e.g. *midagi* [ˈmi.tɑ.ki] 'something'), 33+11+11 in case of a Q2 foot (e.g. *rongile* [ˈroŋ.ki.le] 'train AllSg'), and 333+111+111 in case of a Q3 foot (e.g. *eelmine* [ˈeeː.l.mi.ne] 'previous'). In tetrasyllabic and longer words where successive feet carry secondary stress the labelling was following: 3+1+2+1 for two Q1 feet (e.g. *esimene* [ˈe.si.ˌme.ne] 'first'), 33+11+22+11 for two Q2 feet (e.g. *värviliste* [ˈvær.vi.ˌlis.te] 'coloured one GenPl'), and 3+1+222+111 for a combination of a Q1 and Q3 foot (e.g. *mesilasse* [ˈme.si.ˌlɑsː.se] 'beeyard IllSg').

Stress and quantity levels were identified by a trained phonetician. The measures of duration, F0 mean, F0 standard deviation, and spectral emphasis were extracted using a Praat script. For spectral emphasis the formula SE = L – L_0 was used, where L is the entire intensity of the segment spectrum and L_0 is its intensity in the low band from 0 to 43 % over the F0 median in Hertz (see [16] for details). The data was analysed statistically in R. The mean values of the measures were calculated for each speaker for the factors Style, Stress and Quantity. Three-way Anova tests were used to evaluate the significance of the factors, and post-hoc tests were carried out with Tukey HSD.

3. Results and discussion

The results will be presented separately for each measure studied. For duration the data from male and female speakers was pooled as there is no reason to presume any gender related differences. For F0 and spectral emphasis the analysis was carried out separately for male and female speakers.

3.1. Vowel duration

It can be seen that vowel duration is significant for all factors: Style $F(2, 399) = 97.4$, $p < 0.001$; Stress $F(2, 399) = 69.4$, $p < 0.001$; Quantity $F(2, 399)$ 36.5, $p < 0.001$, and there is a weak interaction of Style and Stress $F(4, 399) = 2.7$, $p < 0.05$, and a strong interaction of Stress and Quantity $F(4, 399) = 41.8$, $p < 0.001$. Post-hoc testing showed that Style is significant on all levels, duration being the longest in word list reading and the shortest in spontaneous speech. For Stress there are diversely different patterns for different quantity levels: in Q2 and Q3, vowel duration is the longest in primary stressed syllables and the shortest in unstressed syllables; but conversely in Q1 the unstressed syllables are the longest and the primary stressed syllables the shortest.

The strong interaction of Stress and Quantity resulting in diversely different stressed-unstressed duration patterns is in line with previous findings from the studies of Estonian quantity system, e.g. [14]. Additionally it can be seen from Figure 1 that there is a large variation of the stressed syllable durations in Q2 and especially in Q3. This is due to the fact that in the present analysis we only considered the foot level quantity but not the stressed vowels being short or long. As pointed out above, the Estonian three-way quantity operates over a disyllabic foot and the main feature of quantity is the relative temporal pattern of stressed and unstressed syllables. The stressed syllables can be filled by various segmental combinations, including a long or an overlong vowel or a combination of a short or a long vowel followed by a short or long coda consonant. Therefore, a more accurate description of stress patterns in Estonian could be obtained by taking the phonological structure into account.

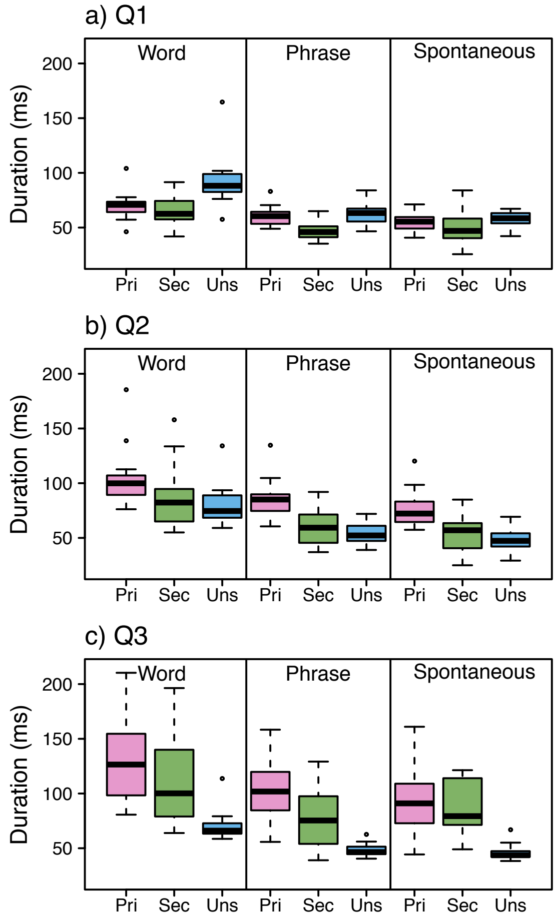

Figure 1: *Duration as a function Style (word list reading, phrase reading, spontaneous) and Stress (primary, secondary, unstressed), grouped by Quantity (Q1, Q2, Q3).*

3.2. F0 mean

For male speakers, the mean pitch is significant for Style $F(2, 151) = 14.9$, $p < 0.001$, and Stress $F(2, 151) = 43.6$, $p < 0.001$. For female speakers, there is only a significant effect of Stress

$F(2, 255) = 17.6$, $p < 0.001$. Post-hoc tests showed that for both male and female speakers, the F0 mean is higher in the primary stressed syllables, but there is no difference between secondary and unstressed vowels. For male speakers, the F0 mean is higher in read speech than in spontaneous speech, but there is no significant difference between the word list reading and phrase reading.

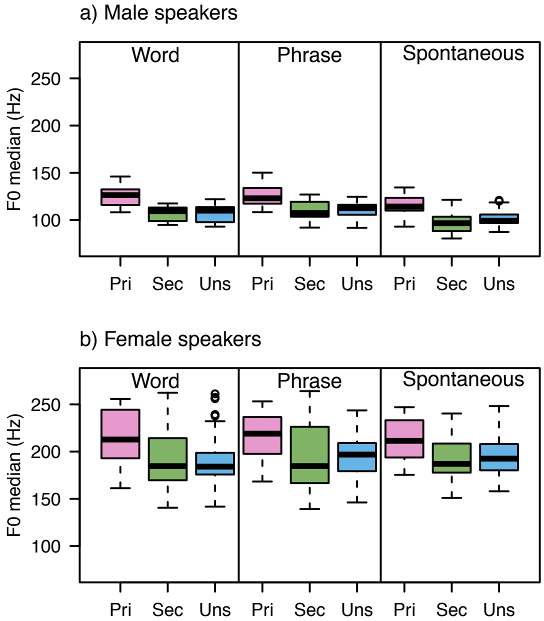

Figure 2: *F0 median as a function Style (word list reading, phrase reading, spontaneous) and Stress (primary, secondary, unstressed).*

Unexpectedly, there was no effect of Quantity on the F0 mean, although it is known from previous studies that there is more pitch movement in Q3 primary stressed syllables.

On the side we noticed a considerable amount of creaky voice in non-initial syllables that were left out of the present analysis. Creakiness usually occurs in Estonian in less prominent parts of speech and can be associated with the lack of prominence. Therefore, some measure of creakiness should be included in the stress model.

3.3. F0 variation

The F0 variation within vowels has different patterns for male and female speakers. For male speakers, there is a significant effect of Stress $F(2, 146) = 12.9$, $p < 0.001$ and a weak interaction of Stress and Quantity $F(4, 146) = 2.4$, $p < 0.05$. Post-hoc testing showed that the F0 variation is greater in primary stressed and unstressed syllables than in secondary stressed syllables. The interaction of Stress and Quantity indicates that the variation of F0 is somewhat different in primary stressed vowels of Q1 vs. Q3 ($p < 0.05$), but in secondary and unstressed syllables the quantity has no effect.

For female speakers, there is a significant effect of Stress $F(2, 249) = 7.4$, $p < 0.001$, and a weak interaction of Style and Stress $F(4, 249) = 2.7$, $p < 0.05$. Post-hoc testing showed that unlike for male speakers the variation of F0 is greater in unstressed vowels, and there is no significant difference between primary and secondary stressed vowels. The post-hoc testing of the interaction of Style and Stress did not reveal any meaningful patterns.

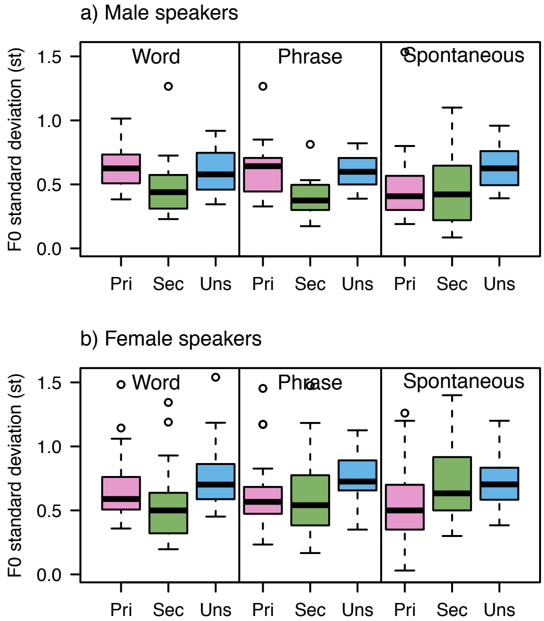

Figure 3: *F0 standard deviation as a function Style (word list reading, phrase reading, spontaneous) and Stress (primary, secondary, unstressed).*

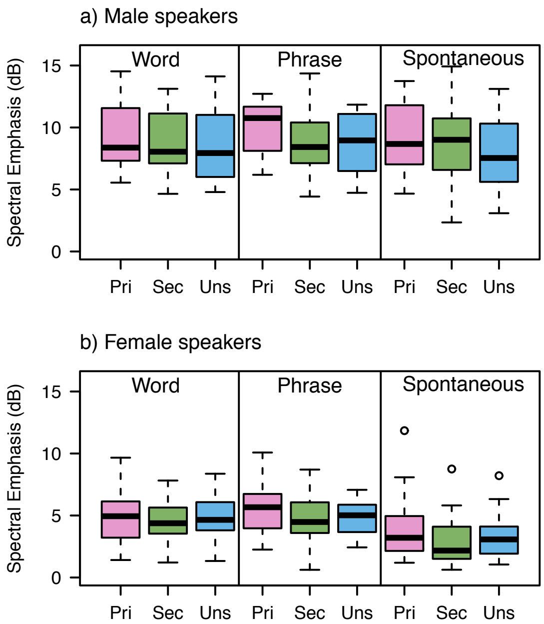

Figure 4: *Spectral emphasis as a function Style (word list reading, phrase reading, spontaneous) and Stress (primary, secondary, unstressed).*

An interaction between Stress and Quantity was expected and could have been even stronger, because it is known that the F0 in Q1 and Q2 stressed syllables is flat, whereas there is an F0 fall within the Q3 stressed syllable [14]. Also, it is possible that a larger F0 variation in unstressed syllables is linked to the fact that a more marked F0 movement takes place in Q1 and Q2. A greater variation in unstressed syllables was also shown for German [5].

3.4. Spectral emphasis

Spectral emphasis is the weakest measure for Style and Stress in Estonian. Again there is a difference between the genders. None of the tested factors are significant for the male speakers, while the data of the female speakers only shows a significant effect of Style $F(2, 253) = 19.7$, $p < 0.001$. A post-hoc test showed that the spectral emphasis is greater in read as compared to spontaneous speech. Additionally, for female speakers, the effect of Stress closely fails to reach the significance level $F(2, 253) = 2.9$, $p = 0.0595$, the spectral emphasis being somewhat higher in primary stressed vowels than in secondary and unstressed syllables.

4. Conclusions

The application of the methodology of the Word Stress Project yielded new results about the acoustic correlates of Estonian word stress. Four measures were studied: duration, F0 mean, F0 standard deviation, and spectral emphasis.

As predicted, vowel duration turned out to be the most important stress correlate. It was shown to be the strongest for Style and Stress with there being a strong interaction with Quantity. In Q2 and Q3 the stressed vowels were longer than the unstressed ones, but in Q1 the unstressed syllables were the longest. The secondary stressed vowels were always shorter than the primary stressed ones.

The pitch in primary stressed vowels was higher than in secondary and unstressed vowels, while the variation of pitch was high in stressed and unstressed vowels and lower in secondary stressed vowels. The F0 variation could be partly quantity-related, but the nature of the data does not enable to draw a solid conclusion. Additionally, F0 mean was associated with speaking style, although this was only apparent for male speakers whose F0 mean was higher in read than in spontaneous speech.

Spectral emphasis turned out to be the weakest measure of stress for Estonian, being only significant for female speakers, distinguishing read from spontaneous speech.

The results seem to support the hypothesis about there being fewer and weaker stress correlates in Estonian. Since Estonian has a fixed word stress and very clear phonological rules determining the stress patterns, it seems logical that it is not as important to mark stress acoustically as it is in many other languages.

In this study neither the duration of the stressed vowel nor the phonological structure of the stressed syllable was taken into account. Thus, in further work on Estonian word stress, not only the foot level quantity but also the segmental structure of the foot should be factored in.

5. Acknowledgements

The authors are very grateful to the speakers who participated in this study. We would also like to thank the participants of the 2nd Word Stress Workshop for feedback and comments. The study was supported by the Estonian Research Council grant IUT2-37, the targeted financing project SF0050023s09, and the National Program for the Estonian Language Technology Project EKT4.

6. References

[1] A. Eriksson, P. A. Barbosa, and J. Åkesson, "Word stress in Swedish as a function of stress level, word accent and speaking style," in *Nordic Prosody. Proceedings of the XIth conference, Tartu 2012*, E. L. Asu and P. Lippus, Eds. Frankfurt am Main: Peter Lang, 2013, pp. 127–136.

[2] A. Eriksson, P. A. Barbosa, and J. Åkesson, "The acoustics of word stress in Swedish: A function of stress level, speaking style and word accent," in *Proceedings of Interspeech 2013*, 2013, pp. 778–782.

[3] P. A. Barbosa, A. Eriksson, and J. Åkesson, "Cross-linguistic similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese," in *Nordic Prosody. Proceedings of the XIth conference, Tartu 2012*, E. L. Asu and P. Lippus, Eds. Frankfurt am Main: Peter Lang, 2013, pp. 97–106.

[4] P. A. Barbosa, A. Eriksson, and J. Åkesson, "On the robustness of some acoustic parameters for signalling word stress across styles in Brazilian Portuguese," in *Proceedings of Interspeech 2013*, 2013, pp. 282–286.

[5] J. Behrens, "Die Prosodie des Wortakzentes in Abhängigkeit von Akzentlevel und Sprechstil," BA Thesis, Christian-Albrechts-Universität zu Kiel, 2013.

[6] I. Lehiste, "Search for phonetic correlates in Estonian prosody," in *Estonian Prosody: Papers from a Symposium*, I. Lehiste and J. Ross, Eds. Tallinn: Institute of Estonian Language, 1997, pp. 11–35.

[7] I. Lehiste, "Segmental and syllabic quantity in Estonian," in *American Studies in Uralic Linguistics*, T. A. Sebeok, Ed. Bloomington: Indiana University Publications, 1960, pp. 21–82.

[8] M. Hint, *Eesti keele sõnafonoloogia I: Rõhusüsteemi fonoloogia ja morfofonoloogia põhiprobleemid*. Tallinn: Eesti NSV Teaduste Akadeemia, 1973.

[9] M. Hint, *Häälikutest sõnadeni*, 2. ed. Tallinn, 1998.

[10] T.-R. Viitso, "Phonology, morphology and word formation," in *Estonian language*, M. Erelt, Ed. Tallinn: Estonian Academy Publishers, 2003, pp. 9–92.

[11] M. Gordon, "Phonetic correlates of stress and the prosodic hierarchy in Estonian," in *Estonian prosody: Papers from a symposium*, I. Lehiste and J. Ross, Eds. Tallinn: Institute of Estonian Language, 1997, pp. 100–124.

[12] P. Lippus, K. Pajusalu, and P. Teras, "The Temporal Structure of Penta- and Hexasyllabic Words in Estonian," in *Proceedings of the 3rd International Conference Speech Prosody*, R. Hoffmann and H. Mixdorff, Eds. Dresden: TUDpress, 2006, pp. 759–762.

[13] A. Eek and E. Meister, "Quality of standard Estonian vowels in stressed and unstressed syllables of the feet in three distinctive quantity degrees," *Linguistica Uralica*, vol. 34, no. 3, pp. 226–233, 1998.

[14] P. Lippus, E. L. Asu, P. Teras, and T. Tuisk, "Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian," *Journal of Phonetics*, vol. 41, no. 1, pp. 17–28, Jan. 2013.

[15] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. Computer program, 2013.

[16] H. Traunmüller and A. Eriksson, "Acoustic effects of variation in vocal effort by men, women, and children," *The Journal of the Acoustical Society of America*, vol. 107, no. 6, pp. 3438–3451, 2000.