Prosodic Characteristics of Vocalic Hesitations in Comparison with Overlong Vowels in Estonian

Rena Nemoto

Institute of Cybernetics at Tallinn University of Technology, Tallinn, Estonia
rena.nemoto@phon.ioc.ee

Abstract

The goal of this paper is to investigate vocalic hesitations in Estonian and compare them to the related vowels of overlong (Q3) quantity degree. We wonder if there are some language-specific characteristics of hesitations. If yes, which kind of characteristics can be observed in Estonian language? We analyze duration, fundamental frequency (f_0), intensity, and first two formants using 39.5 hours of manually transcribed mono- or dialogue speech from a spontaneous speech corpus. Investigated vocalic hesitations and Q3 vowels are: */ee, ää, aa, õõ, öö/*. The characteristics of hesitations as compared to those of Q3 vowels show that hesitations have longer duration range. Hesitations generally include lower f_0 and intensity values. However, the values vary in terms of vowels. First two formants of hesitations tend to be located at more centralized positions in a vocalic triangle than related Q3 vowels.

Index Terms: vocalic hesitation, Estonian, spontaneous speech, prosody

1. Introduction

Hesitations carry language-specific information. The transcription of vocalic hesitations varies across languages, for example uh/um in American English, *euh* in French or *eh* in Spanish suggesting differences in their perception by native speakers [1]. So we wonder what kind of specific can be found in Estonian language. An early Estonian disfluency study by Hennoste [2] described particles and um-s as initiators of repair in Estonian spontaneous speech. One of the most frequent um in spoken Estonian is *ee*. The um *ee* contained 38% of all um-s in the studied corpus. The other frequent um-s were expressed by *õõ*, *ää*, *mm*, etc. But little study has been done to explore these um-s in Estonian at acoustic level. We are then interested in studying these vocalic hesitations at acoustic level. Hesitations are expressed by lengthened vowels (cf. Figure 1 from [3]) or consonant /mm/, or mixing a vowel with a consonant, etc.

Estonian is a language with word-initial lexical stress in general [4], even though some exceptions like loanwords [5] can be found. And Estonian has a three-way quantity system referred as short (Q1), long (Q2), and overlong (Q3) quantity degree. Most studies of the three-way quantity system concentrate on disyllabic words and compare first and second syllables at foot level: for example (Q1) *sada* /sata/, hundred, singular nominative; (Q2) *saada* /saata/, to send, verb: singular imperative; (Q3) *saada* /saa:ta/, to get, verb: infinitive [5]. The quantity system is complex combining durational and tonal components. Lippus [6] explained the f_0 peak as follows: the peak for Q1 and Q2 is located usually in the second half of the stressed first syllable, but for Q3, it usually appears in the beginning of the stressed syllable vowel.

As overlong (Q3) vowels have longest duration among quantity degrees, we compare prosodic and acoustic behaviors between five vocalic hesitations (lengthened vowels) and its related overlong vowels. In this paper, we address the question as follows: how prosodic features of vocalic hesitations behave from the beginning to the end of a segment in comparison with their related Q3 vowels.

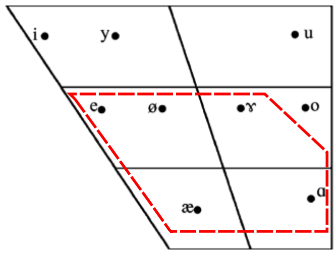

Figure 1: *Estonian vocalic system (vocalic hesitations are in red line).*

2. Corpus and methodology

2.1. Corpus

We used the manually transcribed phonetic corpus of Estonian spontaneous speech of the university of Tartu[1] [6]. Investigated corpus consists of 32 male and 39 female speakers of monologues or dialogues, including 18.5 hours for male and 21 hours for female speakers. The corpus is manually segmented at different levels: phoneme, syllable, word, quantity degree, voice quality, etc.

2.2. Methodology

Fundamental frequency (f_0), intensity, and two first formant (F1 and F2) values were extracted every 5 milliseconds (ms) by using Praat software [7]. Phonemic duration was taken from the manually segmented corpus. f_0, intensity, and first two formant measurements were averaged over segments. Also each phonemic segment was split in three parts (begin, center, end) and measurements of each part were also averaged. Voicing ratio was computed as following: for each segment, the number of voiced (f_0 >0 Hz) frames was divided by the total number of frames. F1 and F2 values were taken only when f_0 values were determined. The target vocalic hesitations and related overlong (Q3) vowels were extracted from the transcribed corpus. Hesitations were annotated by transcriptors. Table 1 gives the

[1] http://www.keel.ut.ee/en/languages-resourceslanguages-resources/phonetic-corpus-estonian-spontaneous-speech

10.21437/SpeechProsody.2014-39

number of occurrences of five vocalic hesitations and related Q3 vowels for male and female speakers. Analyzed Q3 vowels were intra-lexical vowels and vocalic hesitations occurred between silent pauses or words, after a word before a silent pause or inversely (cf. Figure 2).

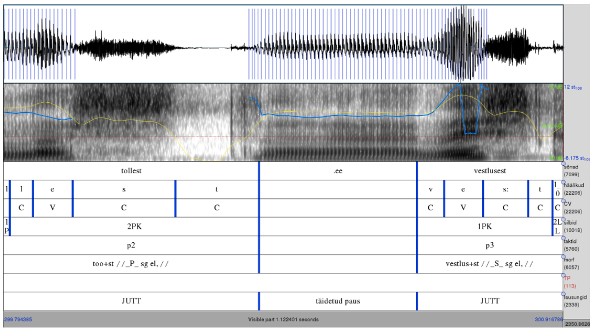

Figure 2: *Example of vocalic hesitation. Vocalic hesitation .ee is located between two words: "tollest (that, singular, elative)#.ee (vocalic hesitation)#vestlusest (conversation, singular, elative)".*

Table 1: *Number of vocalic hesitations and Q3 vowels for male (MS) and female speakers (FS).*

hesit./Q3 vow.	MS		FS	
	hesit.	Q3	hesit.	Q3
.ee/ee: [eeː]	1907	2166	1424	1936
.ää/ää: [ææː]	97	437	102	481
.aa/aa: [ɑɑː]	103	2205	89	2125
.õõ/õõ: [ɤɤː]	608	60	197	41
.öö/öö: [øøː]	68	358	65	359
Total	2,783	5,226	1,877	4,942

3. Acoustic and prosodic analyses

For this study, we compare duration, f_0, intensity, and two first formants (F1 and F2) between five vocalic hesitations and their related overlong (Q3) vowels. We hypothesize that Q3 vowels are intra-lexical vowels, so there are some acoustic and prosodic differences between vocalic hesitations and Q3 vowels.

3.1. Duration

Figure 3 shows duration distribution of five vocalic hesitations (red line) and their related Q3 vowels (black line). Mean hesitation duration reaches 318 milliseconds (ms) for male speakers (standard deviation (SD): 171 ms) and 270 ms for female speakers (SD: 142 ms), whereas mean duration of related Q3 vowels is 154 ms (SD 68 ms) for male and 174 ms (SD 82 ms) for female speakers. We notice that Estonian vocalic hesitations have much longer durations than related Q3 vowels as revealed in the literature like other languages. The profile of vocalic hesitation duration is more enlarged. The difference between hesitations and related Q3 vowels turned out to be statistically significant ($p < .001$) for both male and female speakers by Wilcoxon test using R software [8].

3.2. Fundamental frequency (f_0)

Estonian words in general have a stress with higher f_0 and intensity values on first syllable and these values gradually decrease

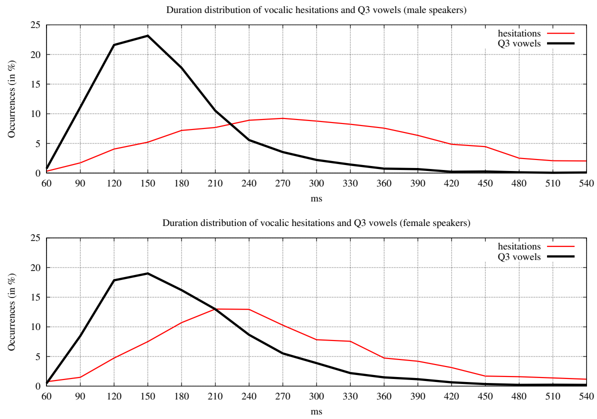

Figure 3: *Duration distribution of vocalic hesitations (in red line) and their related Q3 vowels (in black line) for male (top) and female (bottom) speakers.*

to last syllable. We hypothesize that f_0 contours are more stable for vocalic hesitations while Q3 vowel contours tend to fall within a segment. A segment is divided into three parts (begin, center, end) in order to compare f_0 contour movements of vocalic hesitations from Q3 vowels. Figure 4 presents average f_0 values of begin, center, and end parts over 70% of voicing ratios (to avoid extracting f_0 value errors) for each vocalic hesitation (left) and its related vowel (right). It is noticeable at a glance from Figure 4 that f_0 values are lower for vocalic hesitations than Q3 vowels for both male (top) and female (bottom) speakers. We can also observe that: (i) f_0 values slightly decrease from begin to center parts for both vocalic hesitations and Q3 vowels; (ii) for Q3 vowels, f_0 values continue to fall from center to end parts; (iii) however f_0 contours of vocalic hesitations are stable or they rise from center to end parts (except vocalic hesitation of *.aa* for female speakers where f_0 values continue to fall like Q3 vowels).

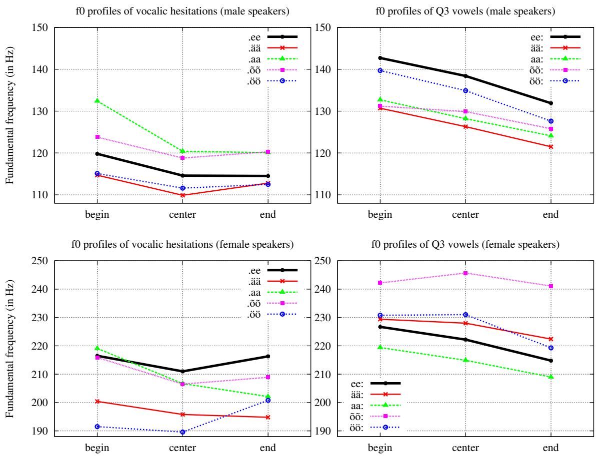

Figure 4: *f_0 contours of vocalic hesitations (left) and their related Q3 vowels (right) for male (top) and female (bottom) speakers.*

To verify whether these f_0 differences between vocalic hesitations and their related Q3 vowels are significant, we conducted statistical tests using Wilcoxon test. First, we compare mean f_0 values of each part (begin, center, end) between vocalic hesitations and their related Q3 vowels. Three pairs (*ee*, *ää*, *öö*) of each part show statistically significant differences ($p < .05$) for both male and female speakers. Especially begin and center parts of these three pairs are strongly significant ($p < .001$) for both gender. However, for the *õõ* pair, only center part is statistically significant ($p < .05$) for both male and female speakers. No significance can be shown in the *aa* pair for both male and female speakers.

Next we compare the difference of f_0 values between begin and center, and between center and end. With respect to the difference of values between each part, the Wilcoxon test shows significant differences in the *ee*, *ää*, and *õõ* pairs for male speakers ($p < .05$), and the *ee* and *õõ* pairs for female speakers ($p < .001$). The difference between begin and center parts of the *aa* pair reveals significant differences for both male and female speakers ($p < .001$). The difference between center and end parts of the *öö* pair are statistically significant for both male and female speakers ($p < .001$) and of the *ää* pair has significant difference for female speakers ($p < .001$).

The *õõ* pair, showing significant difference only in mean center part, has significant difference between begin and center parts, and between center and end parts. The similar phenomenon is observed in the *aa* pair. These suggest that f_0 movements are important to separate vocalic hesitations and its related Q3 vowels.

3.3. Intensity

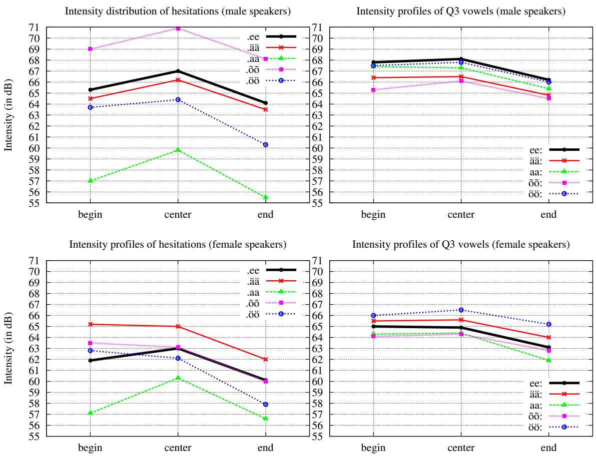

Figure 5: *Intensity distribution of vocalic hesitations (left) and their related Q3 vowels (right) for male (top) and female (bottom) speakers .*

Intensity values are expected to have similar contours as f_0 ones because Estonian words normally have a stress on first syllable. Like f_0, a segment is separated into three parts. Figure 5 (left) shows the mean intensity profiles of begin, center, and end parts of vocalic hesitations for male (top) and female speakers (bottom). For male speakers, intensity values of vocalic hesitations increase from begin to center parts and decrease from center to end parts. Intensity contours of the vocalic hesitations *.ee* and *.aa* for female speakers have also rise-fall contours, whereas the other *.ää*, *.õõ*, and *.öö* hesitation contours decrease from begin to end. However values fall more from center to end than from begin to center. The ranges of each vocalic hesitation values are large for both male and female speakers.

As for the related Q3 vowels (right), mean intensity values are quite stable or values rise slightly from begin to center parts and decrease from center to end parts for both male and female speakers. The range of each Q3 vowel values is not as wide as that of vocalic hesitations for both male and female speakers.

Like f_0, same statistical tests using Wilcoxon are carried out between vocalic hesitations and their related Q3 vowels. The comparison of mean intensity values of each part (begin, center, end) between vocalic hesitations and their related Q3 vowels shows significant differences in the *ee*, *aa*, *õõ*, and *öö* pairs for male speakers ($p < .005$). As for female speakers, the *ee*, *aa*, and *öö* pairs are statistically significant ($p < .001$). Mean intensity values of begin part for the *ää* pair are statistically significant for male speakers, however any statistical significance is not observed in center and end parts. Mean intensity values of begin and center parts do not show any significant difference in the *ää* and *õõ* pairs for female speakers, while significant differences are observed in the end part of the *ää* pair ($p < .001$) and the *õõ* pair ($p < .05$) .

The differences of intensity values between begin and center, and between center and end show significant differences in all pairs for female speakers ($p < .05$). As for male speakers, statistically significant differences are found in the *ee*, *ää*, and *aa* pairs for all pairs ($p < .005$). The difference of intensity values between begin and center part of the *öö* pair are statistically significant ($p < .001$) and also between center and end part of the *õõ* pair ($p < .001$). It is noticeable from statistical tests that not only mean intensity value comparison but also comparison of value difference between three parts are important to separate each characteristics like f_0 contours.

3.4. Formants: F1/F2

As F1 and F2 values of begin and end parts might be influenced by preceding and/or following segments, especially for Q3 vowels, which are in general located within a word, only center part values are taken into account for formant analysis. Center part values seem to give stable values than begin and end parts. Mean F1 and F2 values of center part are illustrated in Figure 6 for male speakers and Figure 7 for female speakers. Vocalic hesitations are presented in red on the top and their related Q3 vowels are in black on the bottom. It is noticeable from two figures that vocalic hesitations (top figures) are superposed each other. However it is observable that vocalic hesitation of *.aa* is less superposed and its ellipse is bigger than other vocalic hesitations for both male and female speakers. On the other hand, Q3 vowels (bottom figures) are distinguishable from each other, although ellipses of *õõ:* and *öö:* are quite superposed. Vocalic hesitation figures show that F1 and F2 positions of these five vocalic hesitations are more centralized to each other than their related Q3 vowels. Mean value of *.ee* is located in more posterior place and that of *.õõ* is more anterior than their related Q3 vowels. Two vocalic hesitations of *.ää* and *.aa* are situated at more closed positions. Concerning the *.öö* hesitation, it seems that mean F1/F2 values are close to those of its related Q3 vowel *öö:*. Vocalic hesitations were manually transcribed, so it might be possible that transcribed vowels and vocalic hesitations were influenced by each transcriptor's perception.

We conducted statistical tests (Wilcoxon test) to verify if these center part values of F1/F2 formants are different between five vocalic hesitations and their related Q3 vowels. Mean center values of F1 for male speakers show significant difference ($p < .001$) except the *öö* pair, whereas mean F2 center values are statistically significant for all pairs ($p < .05$), especially the *ee*, *õõ*, and *öö* pairs ($p < .001$). As for female speakers, statistically

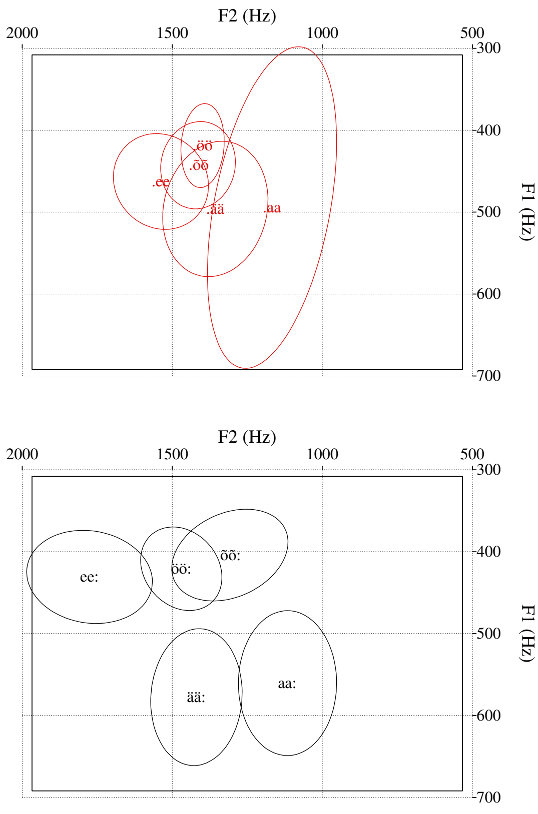

Figure 6: *F1/F2 center position plot of male speaker: hesitations (top, red) and their related Q3 vowels (bottom, black).*

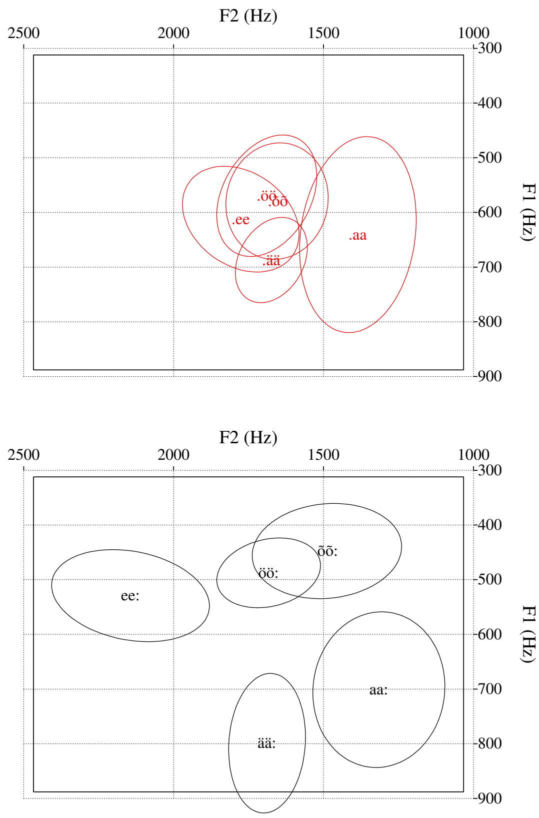

Figure 7: *F1/F2 center position plot of female speaker: hesitations (top, red) and their related Q3 vowels (bottom, black).*

significant differences of F1 values are observed in all pairs ($p < .005$). Concerning F2, no significant difference is found in the *ää* and *öö* pairs, whereas the other pairs show significant difference ($p < .001$). Statistical tests reveal the difference between vocalic hesitations and their related Q3 vowels in all pairs for both F1 and F2 values or for at least one of them.

4. Discussion

This paper described acoustic and prosodic characteristics (duration, fundamental frequency, intensity, and first two formants) of five vocalic hesitations and their related Q3 (overlong) vowels of Estonian language. Investigated five vowels were: *ee*, *ää*, *aa*, *õõ*, and *öö*. Both male and female speakers contained around 5,000 Q3 vowels and 2,800 vocalic hesitations for male speakers and 1,900 for female speakers in 40 hours of the spontaneous speech corpus.

Duration distribution revealed that longer durations were observed in vocalic hesitations than their related Q3 vowels. Duration range was more spread for vocalic hesitations. The f_0 contours of vocalic hesitations behaved differently in comparison with their related Q3 vowels. Lower f_0 values were found in vocalic hesitations. f_0 contours of Q3 vowels tended to decrease from begin to end parts, whereas vocalic hesitation contours decreased begin to center parts and were stable or rising from center to end parts. The results from intensity analysis confirmed different intensity contours between vocalic hesitations and Q3 vowels. The range of vocalic hesitations was bigger than Q3 vowels. Intensity contours of vocalic hesitations raised from begin to center parts and dropped from center to end parts, or decreased from begin to end parts, while Q3 vowel contours showed rise-fall contours. What differentiate vocalic hesitations from Q3 vowels was that intensity values of vocalic hesitations between center and end parts were bigger than those of Q3 vowels. Two first formant results showed that vocalic hesitations were superposed each other and they were more centralized than their related Q3 vowels.

This study concerned general tendencies of vocalic hesitations in comparison with their related Q3 vowels from averaged prosodic and acoustic values using large corpus. We wondered if vocalic hesitations depend on speaker specificities; one speaker might utter hesitations more opened, or other prefers more closed. Hesitations may be influenced by before and/or after phoneme or words. Further studies will include individual speaker aspects and take into account surrounding phonemes or words.

5. Acknowledgements

This research was supported by the European Regional Development Fund (ERDF) through the Estonian Center of Excellence in Computer Science (EXCS) and the Estonian Ministry of Education and Research target-financed research theme No. 0140007s12 to the author. We also thank Einar Meister and Lya Meister for their help and suggestions.

6. References

[1] Vasilescu, I., Nemoto, R. and Adda-Decker, M., “Vocalic hesitations vs vocalic systems: a cross-language comparison”, Proc. of ICPhS, Saarbrücken, 2007.

[2] Hennoste, T., “Repair-initiating particles and um-s in Estonian spontaneous speech”, Proc. of DiSS’05, Aix-en-Provence, pp. 83–88, 2005.

[3] Asu, E. L. and Teras, P., “Estonian”, Journal of the International Phonetic Association, 39, pp. 367–372, 2009.

[4] Asu, E. L. and Nolan, F., “Estonian and English rhythm: a two-dimensional quantification based on syllables and feet”, Speech Prosody, 2006.

[5] Meister, L. and Meister, E., “Perception of the short vs. long phonological category in Estonian by native and non-native listeners”, Journal of Phonetic 39, 212–224, 2012.

[6] Lippus, P., “The acoustic features and perception of the Estonian quantity system, Ph.D. dissertation”, Tartu University, 2011.

[7] Boersma, P. and Weenink, D., Praat: doing phonetics by computer, [Computer program], http://www.praat.org/.

[8] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, âĂŃhttp://www.R-project.or, 2012.