Which kind of hesitations can be found in Estonian spontaneous speech?

Rena Nemoto

Institute of Cybernetics at Tallinn University of Technology, Tallinn, Estonia

Abstract

This paper describes the acoustic characteristics of hesitations in Estonian spontaneous speech. We especially investigate duration, fundamental frequency, and first two formant analyses. Most frequent hesitations can be expressed by lengthened phonemes such as /*ää*/, /*ee*/, /*õõ*/, and /*mm*/. We compare lengthened phoneme hesitations with their related phonemes. The results from our preliminary hesitation study show (i) hesitations have longer duration and its range is spread; (ii) hesitations globally include lower pitch; (iii) hesitation formants are likely to be centralized or posterior and opened in comparison with related phonemes.

Index Terms: hesitation, Estonian, spontaneous speech

1. Introduction

Estonian disfluency has mostly been investigated at spoken language text level such as parsing [1] where the authors paid attention for repairs, repetitions and false starts, but not at acoustic level. In this paper, we investigate hesitations as a preliminary study. In some languages, hesitations are expressed by vowels (such as 'uh/um' or 'er' in English, '*euh*' in French, and '*eh*' in Spanish), and lengthened nasal consonants ('*mm*' in Mandarin) [2]. There is no particular hesitation word in Estonian like English and French. Instead, there are a lot of varieties with lengthened vowels or consonants, or mixing a vowel with a consonant, etc. Estonian has 9 vowels (cf. Figure 1) and 18 consonants. This paper tries to answer the following questions: (i) which vowels or consonants can be used if there is no particular hesitation word? (ii) Which differences can be found between lengthened phoneme hesitations and their related phonemes at acoustic level?

2. Corpus and methodology

For this study, we used the manually transcribed phonetic corpus of Estonian spontaneous speech of the university of Tartu [3]. We used 25 male and 26 female speakers in a corpus of monologues or dialogues, which contains about 15 hours for male and 13 hours for female speakers. Fundamental frequency (f_0) and two first formants (F1 and F2) were extracted every 5 milliseconds (ms) by using Praat software [4]. Measurements were averaged over phonemes. Phonemic duration was taken from the manually segmented corpus.

First we counted occurrences of words transcribed individually in lengthened vowels and consonants. Most frequent lengthened vowels and consonant are presented in Table 1. Figure 1 from [5] shows the Estonian vocalic system with vocalic hesitations expressed in added red line. Frequent vocalic hesitations did not contain close vowels such as [i, y, u]. The most frequent hesitation is *ee* for both male and female speakers with 55% of hesitation occurrences (male: 49% and female 62%). However, each speaker has his/her preference of hesitation use as shown in Table 2. Table 2 compares the hesitation occurrences between three speakers (two male and one female speakers), who contain longer speech duration than other speakers. MS1 is likely to more utter *õõ*, while two other speakers (MS2 and FS3) tend to more often employ *ee*.

Table 1. *Occurrences of hesitation.*

Hesitation	Male	Female	Total
aa	57	52	109
ää	78	102	180
ee	1,029	982	2,011
öö	48	65	113
õõ	572	185	757
mm	297	194	491
Total	2,081	1,580	3,661

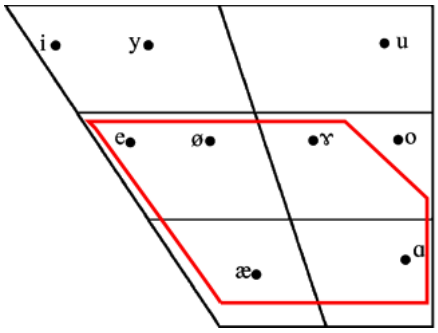

Figure 1: *Estonian vocalic system (vocalic hesitations in red).*

Table 2. *Hesitation occurrences of male speakers (MS1 & MS2) and a female speaker (FS3) with corpus duration in brackets.*

Hesitation	MS1 (45m.)	MS2 (51m.)	FS3 (50m.)
aa	2	6	2
ää	20	23	39
ee	169	311	623
öö	14	4	2
õõ	372	49	47
mm	22	22	3

3. Acoustic analysis

For this study, we compare duration, f_0, and two first formants (F1 and F2) between the hesitations and their related phonemes without considering quantity degree (short, long, overlong).

3.1. Duration

Figure 2 shows duration distribution of hesitations (left), and of related five vowels and one consonant (right). Mean hesitation duration reaches 306 milliseconds (ms) (median: 264 ms, standard deviation: 171 ms), whereas mean related phoneme duration is 78 ms (median: 67 ms, standard deviation: 49 ms). We notice that Estonian hesitations have also much longer durations than related phonemes as revealed

in the literature like other languages. The hesitation duration contour is more spread. The difference between hesitations and related phonemes turned out to be statistically significant ($p < 0.0001$) by Wilcoxon test using R software [6].

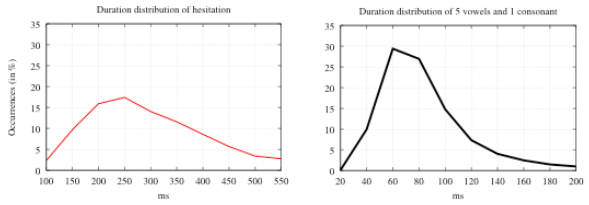

Figure 2: *Duration distribution of hesitations (left: 100–550 ms) and their related phonemes (right: 20–200 ms).*

3.2. Fundamental frequency (f_0)

As each speaker has different pitch, we chose three speakers, who had quite long speech (cf. Table 2), so as to compare f_0 between hesitation and their related vowels. Three frequent vocalic hesitations (*ää*, *ee*, *õõ*) and each related vowel were computed. Table 3 presents mean f_0 in Hz with over 80% of voicing ratios (in order to avoid extracting f_0 value errors) for each vocalic hesitation and its related vowel. Lower f_0 values for vocalic hesitations have been observed for all hesitations and speakers. Statistical analysis using Wilcoxon test showed significant differences of the f_0 values between vocalic hesitations and their related vowels ($p < 0.005$ for all pairs).

Table 3. *Mean f_0 (in Hz) of three speakers (MS1, MS2, FS3) in comparison with three vocalic hesitations and related vowels.*

Hesit./Vowel	MS1	MS2	FS3
ää/ ä (hesit./vow. occ.)	121/135 (20/266)	92/111 (23/327)	235/281 (39/306)
ee/e (hesit./vow. occ.)	126/129 (169/2955)	92/104 (311/1979)	244/264 (623/2034)
õõ/ õ (hesit./vow. occ.)	121/137 (372/405)	94/111 (49/232)	251/284 (47/304)

3.3. Formants

Last we measured two first formants (F1 and F2) through the same three speakers. Figure 3 illustrates F1 and F2 values for three vocalic hesitations (*ää*, *ee*, *õõ*) and their related vowels (*ä*, *e*, *õ*) of two male speakers (MS1 and MS2) and one female speaker (FS3).

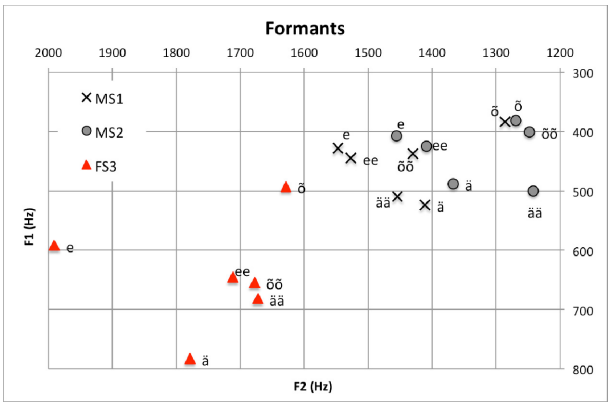

Figure 3: *F1/F2 mean values (in Hz) for three vocalic hesitations (ää, ee, õõ) and their related vowels of three speakers (MS1, MS2, and FS3).*

In comparison with vocalic hesitations and their related vowels, we can observe that MS1 has three vocalic hesitations centralizing to each other especially *õõ*, while MS2 globally contains more opened and posterior vocalic hesitations than the vowels. Three vocalic hesitations of FS3 are also centralized and very close to each other. We conducted statistical tests to verify if these formants are different between vocalic hesitations and related phonemes. As for FS3, both F1 and F2 of all pairs were significantly different ($p < 0.001$). There is no difference of F1 between the *ää*/*ä* pair for MS1 and MS2. However, the other pairs of F1 were significantly different ($p < 0.05$). As for F2, no significant difference is found in the pair *ee*/*e* for MS1 and the pair *õõ*/*õ* for MS2, whereas the others showed significant difference ($p < 0.01$).

4. Discussion

In this paper, we aimed at exploring the preliminary hesitation study of Estonian language. Our study focused on especially acoustic characteristics (duration, f_0, two first formants) of Estonian hesitations in a spontaneous speech corpus. As characteristics of hesitations in Estonian, we found that lengthened vowels such as /*aa*/, /*ää* /, /*ee*/, /*õõ* /, /*öö* /, and a lengthened consonant /mm/ were mostly used. The duration comparison between lengthened phoneme hesitations and their related phonemes showed that hesitations include longer duration and the duration range is spread. The result from f_0 analysis of three speakers revealed that vocalic hesitations have globally lower values. Two first formants of vocalic hesitations tended to be more centralized or posterior and opened than related phonemes. However the degree of centralization and aperture is dependent on speakers. In the future, we will study in-depth this point with more speakers.

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.

6. References

[1] K. Müürisep and H. Nigol, "Shallow Parsing of Transcribed Speech of Estonian and Disfluency Detection", *Human Language Technology*, Springer Verlag, pp. 165–177, 2009.

[2] I. Vasilescu et al., "Vocalic hesitations vs vocalic systems: a cross-language comparison", *Proc. of ICPhS*, Saarbrücken, 2007.

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

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

[5] E.L. Asu and P. Teras, "Estonian", *Journal of the International Phonetic Association* 39, pp. 367–372, 2009.

[6] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.or, 2012.