The tonal component of Estonian quantity in native and non-native perception

Pärtel Lippusa,*, Karl Pajusalu a, Jüri Allikb

a Institute of Estonian and General Linguistics, University of Tartu, Ülikooli 18, Tartu 50090, Estonia
b Institute of Psychology, University of Tartu, Tiigi 78, Tartu 50410, Estonia

Received 9 December 2008; received in revised form 3 July 2009; accepted 6 July 2009

Abstract

This paper focuses on the role of the pitch cue in the perception of Estonian quantity degrees. The significance of the tonal component is investigated through comparison of native vs. learned identification of the quantities. The study reports on perception tests with manipulated natural speech stimuli given to Estonian native (L1) listeners and non-native (L2) listeners with different language backgrounds. Our earlier results [Lippus, P., Pajusalu, K., & Allik, J. (2007). The tonal component in perception of the Estonian quantity. In J. Trouvain, & Barry, W. J. (Eds.), The proceedings of the 16th international congress of phonetic sciences: 16th international congress of phonetic sciences, Saarbrücken, Germany, 6–10 August 2007 (pp. 1049–1052)] showed that native Estonian listeners use pitch as an important cue for perceiving the overlong quantity (Q3), but the cue plays no significant role for L2 listeners. In this study we analyze more closely the effect of the listeners’ native language on the perception of Estonian quantities. Finnish and Russian L1 listeners successfully learn the Estonian three-way quantity distinction without considering the pitch cue. Latvian L1 listeners show some confusion between the long and overlong quantity degree, which could be explained by reference to the tonal and temporal contrasts in their native language.

1. Introduction

A central feature of Estonian word prosody is the three-way quantity distinction that is manifested as complex prosodic structures which combine durational and tonal components. The distinction of short (Q1), long (Q2), and overlong (Q3) quantity degrees can be realized in a disyllabic foot by the primary-stressed syllable vowel (V1) (e.g. [satu] ‘hundred’-[su:tta] ‘to send, sg imperative’-[sa:tta] ‘to get’), syllable-medial consonants (C2) (e.g. [kutta] ‘slingshot’-[kutta] ‘to cover, sg imperative’-[kutta] ‘to cover’), or both (e.g. [sättte] ‘to get, pl 2nd pers’-[sättte] ‘broadcast, sg gen’). There is no phonological length opposition in the unstressed syllable, but due to foot isochrony the duration of the unstressed vowel in the second syllable (V2) is longest in Q1 and shortest in Q3. Although the main carrier of the quantity is the vowel in the first syllable or the syllable-final consonant, the opposition is better described by the relationship between segmental and syllabic duration in a disyllabic foot (Asu, Lippus, Teras, & Tuisk, in press; Eek & Meister, 2003, 2004; Lehiste, 1960, 1997, 2003; Liiv, 1961; Traunmüller & Krull, 2003).

Starting from Lehiste’s and Liiv’s work in the 1960s the Estonian quantities have mainly been described by comparing the duration of the first and second syllable of a foot. Since the duration of the syllable-initial consonant is not contrastive, the syllable onset is usually excluded from comparison, and only the duration ratio of syllable rhymes is computed. The non-initial syllable begins with a single consonant; in words with a geminate consonant on the syllable boundary, the geminate is divided so that the second part belonging to the second syllable onset has approximately the same duration as the corresponding singleton consonant. The typical S1/S2 ratio is 2/3 in a Q1 foot, 3/2 in a Q2 foot, and 2/1 in a Q3 foot (Lehiste, 1997). Similar results are also reported by others (e.g. Eek & Meister, 1997; Krull, 1991, 1992; Liiv, 1961).
Besides the temporal features, pitch has also been shown to play an important role in the quantity distinction. In Q1 and Q2 there is a step-down in the f0 contour between the end of the first syllable nucleus and the beginning of the second syllable, while Q3 is associated with a fall early during the first syllable (Asu, 2004; Asu et al., in press; Lehiste, 1960; Liiv, 1961; Remmel, 1975).

Previous perception studies (Eek, 1980; Lehiste, 1970) have shown that the pitch cue is crucial for distinguishing Q2 and Q3. Lehiste and Danforth (1977) present a hierarchy of phonetic cues for the perception of the Estonian quantities according to which the pitch cue is the second most important after the duration of V1. On the basis of her perception tests Lehiste (1997) concludes that the quantity opposition is binary: the syllable ratios discriminate short from long, but for the discrimination of long and overlong, the pitch cue is vital.

The syllable duration ratio is a good characteristic for describing the production of Estonian quantity, but it has been criticized from the point of view of perception. Traunmüller and Krull (2003) found a number of reasons why comparing the S1/S2 duration ratio might not work for perception. In Lehiste and Danforth’s hierarchy the most important cue is the V1 duration, but Traunmüller and Krull argue that if the quantity degree is identified by the S1/S2 ratio, the second syllable should have the same perceptual weight as the first syllable. Furthermore, if the consonant in the position of C2 is a geminate, the syllable boundary should lie somewhere within the geminate, but there is no objective way to find it. The importance of pitch as the main cue for distinguishing Q3 from Q2 has also been questioned, as in many cases there is a voiceless consonant in the S1 rhyme. Instead of the S1/S2 ratio in combination with the pitch contour, Traunmüller and Krull (2003) suggest that the quantity is identified by comparing the V1 duration with the weighted sum of segment durations within the foot. The effect of the neighbouring segment duration has also been tested by Eek and Meister (2003, 2004), who propose a model for production and perception of the quantity degrees where the durations of the S1 nucleus, S1 coda (in case of Q2 and Q3), and S2 nucleus are compared, while the syllable onsets indicate the local speaking rate. For describing production, this model may be slightly more precise, but one still has to find the imaginary syllable boundary if the consonant in C2 position is a geminate, and also postulate the imaginary nucleus/coda boundary in the case of long V1.

Related to the temporal structure of the feet, vowel quality has been shown to vary in connection with quantity. In the stressed syllable, the vowels in Q3 feet are the most peripheral (while longest in duration) and vowels in Q1 feet are the most centralised. Showing greater variation in the unstressed syllable, the quantity degrees affect the vowel quality in the opposite direction: the vowels of Q1 feet are least reduced (and longest in duration) and the vowels of Q3 feet, most centralised (Eek & Meister, 1998). Likewise, changes in segment duration have been shown to affect the perception of vowel quality (Meister, Meister, & Werner, 2008). The perception of vowel quality as a cue to the Estonian quantity opposition has not been investigated.

The perception of Estonian quantity degrees by second language (L2) listeners of Estonian has mainly been studied in connection with non-native accent in Estonian. Pajusalu (1994) has investigated how native Finnish listeners perceive the quantity degrees in Estonian. His results show that native Finnish listeners attribute more phonological weight to the length of the unstressed syllable, similarly to Finnish, where there is a phonological opposition of long and short duration in unstressed syllables, and they perceive the Estonian words with Q2 similarly to Finnish words with long first and second syllables.

In Lippus, Pajusalu, and Allik (2007) we reported the results of a group of 9 native vs. 9 near-native Estonian listeners with different first languages (L1). The results of the Estonian L1 listeners showed that in the case of vowel quantity, the test subjects failed to perceive Q3 if the pitch cue was that of Q1 or Q2, but distinguished between all the quantity degrees successfully if the consonant quantity was carried by a voiceless stop. In the case of the Estonian L2 listeners the various f0 contours had no effect, and all the quantity levels were perceived correctly with all the sets of stimuli. The L2 listeners differed in that their crossover points in all sets were not as clear as those of the Estonian L1 listeners. In Lippus and Pajusalu (in press) we observed a larger group of Estonian L1 listeners. Within the 35 test subjects, variation due to dialectal background appeared. The subjects from western and central Estonia also used the pitch cue to decide between Q2 and Q3, while the subjects from eastern and southern Estonia based their decision primarily on temporal cues. We propose that the perception of Estonian quantity degrees is based on the interaction of various cues: the temporal structure may be enough in most cases, but the pitch cue (and additionally other cues like vowel quality and context) is also taken into account. If the different cues are conflicting, the identification of quantity can be disturbed.

In this paper we take a closer look at non-native Estonian listeners with different first languages. Using the same test set-up as in Lippus et al. (2007), groups of Finnish, Russian, and Latvian L1 listeners were tested. On the one hand, these languages are spoken in areas neighbouring Estonia, and they have all had historical contacts with Estonian; on the other hand, these languages have different prosodic systems. In Finnish, the stress is fixed on the word-initial syllable, and the opposition is only between short and long segments and can occur in most positions in the word. Stress is realised by durational patterns; pitch variation and vowel reduction are not significant (Suomi, Toivanen, & Ylitalo, 2008). In Russian, stress is not fixed, and it is marked with vowel length and reduction (Kodzasov & Krivnova, 2001). In Latvian, fixed stress on the word-initial syllable is associated with pitch, and an opposition of short and long vowels can occur in...
any position in the word. While Finnish and Russian do not exhibit any significant word-level tonal alternations, Latvian is characterised by three lexical tones or pitch accents: level, falling, and broken (similar to Danish stød, realized by laryngealization or a glottal stop; Bond, Stockmal, & Markus, 2006; Karins, 1996). It is known from many L2 studies that L2 perception and production are largely based on the L1, and if the features used to mark categories in L2 are also used in L1, more effective acquisition of L2 can be expected. On the other hand different use of the same features in L1 and L2 can obstruct the acquisition of L2 categories, while categories and features of L2 that are unfamiliar in L1 are learned more successfully (e.g. Aoyama & Guion, 2007; Best & Tyler, 2007; Cutler, Mehler, Norris, & Segui, 1989; McAllister, 2007). We would therefore expect the Latvian L1 listeners to benefit from both pitch and duration contrasts and perceive Estonian quantity similarly to the Estonian L1 listeners, while the Finnish L1 and Russian L1 listeners would be expected to ignore the pitch cue and show less similarity to the Estonian L1 listeners.

2. Materials and methods

The perception test involved assigning quantity degrees to stimuli synthesized from natural words with Q1, Q2, and Q3. The stimuli were created using resynthesis in Praat (Boersma & Weenink, 2007). For the test stimuli two quantity triplets were used as base words: one for the vowel quantity where the quantity distinction is carried by V1 followed by a short consonant (sada [sata] ‘hundred’, saada [satu:a] ‘to send, sg imperative’, saada [satu:a] ‘to get’), and the other for the consonant quantity where the quantity distinction is carried by the intervocalic consonant after a short V1 (kada [kata] ‘slingshot’, kata [katta] ‘to cover, sg imperative’, katata [katata] ‘to cover’). The words were read by a male speaker in a nonsense carrier phrase ‘tule … saama’ (‘come … to get’) with the test word in a focal position preceded and followed by a disyllabic word. From each of the six words a set of nine stimuli was created by manipulating the duration of either the first vowel (V1) or the intervocalic consonant (C2); a total of 54 different stimuli were created. The stimuli were synthesized so that for the first stimuli the syllable duration ratio would be smaller than 2/3 and for the last larger than 2/1.1 The duration of only one sound in the word was changed, in nine steps of 30 ms, starting from 50 ms to 290 ms. The locations of the pitch turning points remained proportionally unchanged. The pitch curves of the original words are presented in Fig. 1.

![](image.png)

Fig. 1. (Left panel) The pitch curves in the words [sata] (Q1; solid line), [satu:a] (Q2; dashed line), [satu:a] (Q3; dotted line). (Right panel) The pitch curves in the words [kata] (Q1; solid line), [katta] (Q2; dashed line), [katata] (Q3; dotted line).

A three-alternative forced-choice perception experiment was carried out in Praat. The nine stimuli from each base word were arranged as one sub-test and presented to the listeners without the carrier phrase with 10 repetitions in random order (i.e. 6 × 9 × 10 = 540 stimuli in total). Each sub-test was preceded by an exercise where the nine stimuli were presented without repetitions. The subjects heard resynthesized words and had to decide whether they heard a Q1, a Q2, or a Q3 word. They were instructed to think about the meaning of each word and click a button on the computer screen, labelled [1], [2], and [3] accordingly. The subjects were allowed to have a small break between each sub-test. No repeated listening option was available.

The test subjects form four groups: 10 Estonian L1 listeners, 6 Finnish L1 listeners, 6 Russian L1 listeners, and 9 Latvian L1 listeners (the age and sex of the subjects are given in Table 1). The subjects in the Finnish L1 listeners

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
<th>Mean age</th>
<th>Min. age</th>
<th>Max. age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estonian L1</td>
<td>6</td>
<td>4</td>
<td>32</td>
<td>18</td>
<td>43</td>
</tr>
<tr>
<td>Finnish L1</td>
<td>–</td>
<td>6</td>
<td>34</td>
<td>23</td>
<td>46</td>
</tr>
<tr>
<td>Russian L1</td>
<td>5</td>
<td>1</td>
<td>26</td>
<td>21</td>
<td>38</td>
</tr>
<tr>
<td>Latvian L1</td>
<td>8</td>
<td>1</td>
<td>30</td>
<td>21</td>
<td>44</td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>12</td>
<td>31</td>
<td>18</td>
<td>46</td>
</tr>
</tbody>
</table>

1For checking the S1/S2 duration ratio of the stimuli, in case of V1 lengthening the duration ratio of V1/V2 was computed, excluding the syllable onset consonants; in case of C2 lengthening, the duration of C2 of the Q1 word [kata] was taken as the duration of S2 onset in a geminate consonant and the duration ratio of (V1 + C2–C2[kata])/V2 was calculated.
group had learned Estonian as adults. By the time of participation in the test they had all lived in Estonia for at least a year and used Estonian as their home and/or work language. Four subjects in the Russian L1 listeners group had learned Estonian in Russian-taught schools in Estonia, and two subjects had learned Estonian as adults upon moving to Estonia to study at the university. They all lived permanently in Estonia and used Estonian as their work language. The subjects in the Latvian L1 listeners group had learned Estonian at the university in Estonia, but after graduation most of them moved back to Latvia. They spoke Latvian as their home and work language, but also used Estonian on a daily basis (as teachers or tour guides, etc.) and frequently visited Estonia, where they spoke Estonian. The language skills of the Estonian L2 subjects were not tested, but they spoke Estonian equally fluently, albeit with a slight accent.

3. Results

The results of the four groups were tested with logistic regression analysis using the stepwise procedure for selecting the significant factors from V1, C2, and V2 durations and the pitch contour (assuming that pitch contour is the main feature that is left of the original base word after manipulating the duration). Separate models of Q1 responses and Q3 responses were created for all the groups and for the stimuli with vowel lengthening and consonant lengthening. As the main difference between the sub-tests and the subject groups lies in the Q3 responses, only the models of Q3 responses are discussed here.

The logistic regression analyses of the Q3 responses to the stimuli with vowel lengthening are presented in Table 2. Fig. 2 presents both the actual responses to the stimuli with vowel lengthening and logistic regression lines.

The results of the Estonian L1 listeners show that in the case of vowel lengthening, the pitch cue is important for deciding between Q1 and Q2 vs. Q3. The stepwise logistic regression model of Q3 responses shows the main effect of V1 duration and interaction of the V1 duration and the pitch contour as significant (see Table 2). In the first sub-test with the Q1 base word [s<sub>t</sub>] with flat pitch in the first syllable, Q1 was perceived when the S1/S2 ratio was 0.4–0.9 and Q2 was perceived when the S1/S2 was 1.2–1.6, but when the S1/S2 was 1.9–2.4, identification of Q3 vs. Q2 was near chance level (Fig. 2a). In the second sub-test, where the stimuli were created from the Q2 base word [s<sub>t</sub>] with flat pitch, the distinction between quantities was similar: Q1 was perceived when the S1/S2 was 0.5–0.8, and Q2 when the S1/S2 was 1.4–2.0, but when the S1/S2 was 2.3–2.9 the identification of Q3 vs. Q2 was near chance level (Fig. 2b).

In the third sub-test with the Q3 base word [s<sub>u</sub>u<sub>r</sub>्] with falling pitch in the first syllable, all the quantity degrees were perceived according to temporal structure: Q1 when S1/S2 was 0.7, Q2 when S1/S2 was 1.0–1.8, and Q3 when S1/S2 was 2.2–3.8 (Fig. 2c). In the logistic regression model only the interaction of V1 duration and pitch contour of Q3 had a significant effect for Q3 responses, but Q2 pitch contour vs. Q1 contour was not significant (see Table 2). Thus, for the Estonian L1 group the tonal cue was important for perceiving Q3.

### Table 2

Logistic regression analysis of Q3 responses for the stimuli with vowel lengthening using the stepwise procedure for selecting the significant factors.

|                     | Estimate | Std. error | z value | Pr (>|z|) |
|---------------------|----------|------------|---------|----------|
| **Estonian L1 listeners** |          |            |         |          |
| Intercept           | -9.294   | 0.778      | -11.939 | <0.001***|
| V1 duration         | 0.034    | 0.003      | 11.077  | <0.001***|
| Pitch Q2            | 0.437    | 1.054      | 0.415   | 0.678    |
| Pitch Q3            | 1.199    | 0.974      | 1.231   | 0.218    |
| V1 dur: pitch Q2    | 0.001    | 0.004      | -0.230  | 0.818    |
| V1 dur: pitch Q3    | 0.019    | 0.005      | 3.857   | <0.001***|
| **Finnish L1 listeners** |          |            |         |          |
| Intercept           | -12.073  | 0.855      | -14.113 | <0.001***|
| V1 duration         | 0.041    | 0.002      | 19.797  | <0.001***|
| C2 duration         | 0.038    | 0.007      | 5.284   | <0.001***|
| **Russian L1 listeners** |          |            |         |          |
| Intercept           | -9.720   | 0.518      | -18.754 | <0.001***|
| V1 duration         | 0.044    | 0.002      | 19.406  | <0.001***|
| Pitch Q2            | -0.182   | 0.213      | -0.852  | 0.394    |
| Pitch Q3            | 0.700    | 0.215      | 3.252   | 0.001**  |
| **Latvian L1 listeners** |          |            |         |          |
| Intercept           | -4.8E+00 | 1.80E+00   | -2.661  | 0.008**  |
| V1 duration         | 8.08E-03 | 8.29E-03   | 0.976   | 0.329    |
| C2 duration         | -1.46E-02| 2.04E-02   | -0.718  | 0.473    |
| V1 dur: C2 dur      | 2.20E-04 | 9.48E-05   | 2.316   | 0.021*   |

*** stands for $p<0.001$, ** for $p<0.01$, and * for $p<0.05$. 
In the case of vowel lengthening, the groups of Finnish L1 listeners (Figs. 2d–f) and Russian L1 listeners (Figs. 2g–i) perceived all quantity degrees according to the temporal structure of the stimuli, and the pitch cue had no effect. Logistic regression analysis of Q3 responses of the Russian and Finnish L1 listener groups showed significant effects of V1 duration and also of the Q3 pitch contour, though the responses of the Russian and Finnish L1 listener groups look extremely similar. The results were also tested with Pearson’s $\chi^2$ test for the equality of proportions of the responses between the subject groups. There were no significant differences between the Finnish L1 and Russian L1 groups: in sub-test 1 $\chi^2(2) = 1.114; p = 0.573$, in sub-test 2 $\chi^2(2) = 5.797; p = 0.055$, and in sub-test 3 $\chi^2(2) = 0.145; p = 0.930$. Differences between the Estonian L1 group on the one hand and the Finnish L1 and Russian L1 groups on the other hand are significant: in sub-test 1 $\chi^2(4) = 105.777; p < 0.001$, in sub-test 2 $\chi^2(4) = 105.106; p < 0.001$, and in sub-test 3 $\chi^2(4) = 59.34; p = 0.001$.

The results of the Latvian group are presented in Fig. 2j–l. In the Latvian group, there is much deviation in responses to the stimuli that are not judged as Q1 and the stimuli get barely more than 70% of single quantity degree judgements. In all sub-tests the shorter stimuli with the S1/S2 ratio less than 1 were judged as Q1, similarly to the Estonian L1 listeners. In sub-tests with vowel lengthening, the results of the Latvian group are similar to the results of the Estonian group in that in sub-tests 1 and 2 where the pitch was flat in the first syllable, the quantity identification level was rather low in most cases where the S1/S2 ratio was greater than 2, but in sub-test 3 where the pitch was falling in the first syllable, the perception of Q3 was high. The logistic regression analysis found quite low significance of the interaction effect of V1 duration and C2 duration to the Q3 responses, but no effect of pitch contour was found (see Table 2). Even though the identification of Q3 seems to be similar between the Estonian L1 group and the Latvian L1 group, the response pattern is significantly different (in sub-test 1 $\chi^2(2) = 56.025; p < 0.001$, in sub-test 2 $\chi^2(2) = 51.359; p < 0.001$, and in sub-test 3 $\chi^2(2) = 57.409; p < 0.001$) and the logistic regression analysis refutes the significance of tonal characteristics for the Latvian L1 group.

In the case of consonant lengthening, the pitch contour was interrupted by the voiceless stop in the C2 position. The pitch cue was not available and all the quantity degrees were perceived according to the temporal structure of the stimuli in all groups. Logistic regression analyses of the Q3 responses to the stimuli with consonant lengthening are presented in Table 3. In Fig. 3 both actual responses to the stimuli with consonant lengthening and logistic regression lines are presented.

![Figure 2](image-url)
Table 3
Logistic regression analysis of Q3 responses for the stimuli with consonant lengthening using the stepwise procedure for selecting the significant factors.

| Estimator     | Estimate | Std. error | z value | Pr(>|z|) |
|---------------|----------|------------|---------|----------|
| Intercept     | -6.394   | 0.369      | -17.330 | <0.001***|
| C2 duration   | 0.050    | 0.002      | 24.710  | <0.001***|
| V2 duration   | -0.033   | 0.003      | -11.680 | <0.001***|

**Estonian L1 listeners**
- Intercept: -6.471, Std. error: 0.730, z value: -8.870, Pr(>|z|): <0.001***
- C2 duration: 0.039, Std. error: 0.002, z value: 19.436, Pr(>|z|): <0.001***
- V1 duration: 0.009, Std. error: 0.001, z value: 19.436, Pr(>|z|): <0.001***

**Finnish L1 listeners**
- Intercept: -5.385, Std. error: 0.698, z value: -7.713, Pr(>|z|): <0.001***
- C2 duration: 0.040, Std. error: 0.002, z value: 20.267, Pr(>|z|): <0.001***
- V1 duration: 0.009, Std. error: 0.001, z value: 19.436, Pr(>|z|): <0.001***

**Russian L1 listeners**
- Intercept: -3.143, Std. error: 0.423, z value: -7.427, Pr(>|z|): <0.001***
- C2 duration: 0.019, Std. error: 0.001, z value: 23.464, Pr(>|z|): <0.001***
- V1 duration: 0.006, Std. error: 0.002, z value: 19.436, Pr(>|z|): <0.001***

**Latvian L1 listeners**
- Intercept: -3.143, Std. error: 0.423, z value: -7.427, Pr(>|z|): <0.001***
- C2 duration: 0.019, Std. error: 0.001, z value: 23.464, Pr(>|z|): <0.001***
- V1 duration: 0.006, Std. error: 0.002, z value: 19.436, Pr(>|z|): <0.001***

*** stands for $p<0.001$, ** for $p<0.01$, and * for $p<0.05$.

Fig. 3. The judgments of the Estonian L1 (left column), Finnish L1 (mid left), Russian L1 (mid right), and Latvian L1 listeners (right column) to the stimuli with manipulated C2 duration. The sub-tests with the Q1 base word are in the first row, the sub-tests with the Q2 base word in the second row, and the sub-tests with the Q3 base word in the third row. The circles represent the actual Q1 responses, the triangles the Q2 responses, and the crosses the Q3 responses. The regression lines represent the approximation of the logistic regression analysis of Q1 and Q3 responses: $p_{Q1}$ with the solid line, $1-(p_{Q1} + p_{Q3})$ with the dashed line, and $p_{Q3}$ with the dotted line.

In the Estonian L1 group, the logistic regression analysis found the C2 and V2 duration to be significant factors for Q3 responses (see Table 3). In sub-test 4 with the Q1 base word [kutu], Q1 was perceived with the S1/S2 ratio 0.5–0.7, Q2 with S1/S2 1.0–1.5, and Q3 with S1/S2 2.0–2.5 (Fig. 3b). In sub-test 5 with the Q2 base word with a short geminate [kutu], Q1 was perceived with the S1/S2 ratio 0.5–0.7, Q2 with S1/S2 1.0–1.5, and Q3 with S1/S2 2.0–2.5 (Fig. 3b). In sub-test 6 with the Q3 base word with a long geminate [kutu], Q1 was perceived with the S1/S2 ratio 1.1, Q2 with S1/S2 2.0–2.5, and Q3 with S1/S2 3.5–4.9 (Fig. 3c). The results show that in the case of short V1 and the
lengthening of the following voiceless consonant, the tonal contrast does not appear.

The logistic regression analysis of Q3 responses in the case of consonant lengthening found a significant effect of V1 and C2 duration in all three Estonian L2 listener groups (see Table 3). Again the response pattern of the Finnish L1 and Russian L1 groups are similar (Fig. 3d–f and g–i). The Finnish L1 and Russian L1 listeners perceived all the quantity degrees according to the durational pattern of the stimuli in all sub-tests. In the Finnish and Russian L1 listener groups the crossovers between the perceived quantity degrees in all sub-tests were not as sharp as in the Estonian L1 group, but the crossover points were linked with the same stimuli as for the native Estonian listeners.

In the case of consonant lengthening, in the Latvian L1 group all quantity degrees were perceived according to the temporal structure of the stimuli, but differently from the other groups; no stimulus got more than 75% single quantity degree judgements. Therefore, the absence of tonal information disturbs the perception of the Latvian L1 listeners.

5. Conclusions

The experiment confirms the importance of the pitch cue for Estonian L1 listeners in the perception of Estonian quantity degrees. In order to successfully perceive Q3, a falling pitch in the first syllable is needed, in addition to the temporal characteristics. Flat pitch in the first syllable.
triggers the perception of Q2 even if the temporal structure is that of Q3. If there is no pitch cue, the temporal structure is sufficient for the perception of all quantity degrees. Rather than positing only duration for distinguishing Q1 from Q2 and only pitch for distinguishing Q2 from Q3, we suggest that there is a trade-off between different cues including the temporal structure, pitch, and vowel quality, not to mention the context in natural speech situations.

The experiment also shows that Estonian L2 listeners are affected by the prosodic system of their first language. The Latvian L1 listeners, whose first language uses both temporal and tonal features, seem to rely more on tone for distinguishing between long and overlong quantities than Estonian L1 listeners. Since in their L1 they use the two features for different categories (pitch for signalling tones and duration for signalling quantity), we assume they also treat Estonian quantity as two categories, using pitch mainly to identify Q2 vs. Q3, but the different use of these features in Estonian disturbs their decisions. In the case of consonant lengthening, for Latvians a missing pitch cue can mask the distinction between Q2 and Q3. Unlike the Latvian L1 listeners, Finnish and Russian L1 listeners successfully learn to identify the Estonian quantity degrees by the temporal structure, since neither Finnish nor Russian uses tonal features on the word-prosody level.

Acknowledgements

The present research was partly supported by the Estonian Science Foundation Grant no. 7904. We would like to thank all our test subjects. We are also very grateful to Arvo Eek, Jaan Ross, and Eva Liina Asu for their constructive comments, to Associate Editor Ocke-Schwen Bohn, and to two anonymous reviewers for numerous helpful suggestions, and to Virve-Anneli Vihman for language editing this paper.

References


