Journal of Phonetics 39 (2011) 212–224

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Perception of the short vs. long phonological category in Estonian by native and non-native listeners

Lya Meister, Einar Meister *

Laboratory of Phonetics and Speech Technology, Institute of Cybernetics at Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia

ARTICLE INFO

Article history:
Received 8 June 2010
Received in revised form
17 January 2011
Accepted 24 January 2011

ABSTRACT

This paper studies the perception of Estonian duration-based phonological oppositions by native Estonians and non-native speakers with Russian-language background. The short/long category boundary was examined by varying the duration of a vowel in three contexts involving isolated vowels (V vs. VV), one-syllable nonsense words (CVC vs. CVVC), and two-syllable real words (CVCV vs. CVVCV). Since vowel duration serves to distinguish lexical minimal pairs in Estonian but not in Russian, L1 and L2 subjects are expected to employ different perceptual strategies in a short/long categorization task. In particular, location and width of category boundaries as well as consistency of categorization are likely to vary between the groups. The results showed that L2 subjects were quite successful in distinguishing the Estonian short/long categories despite the non-categorical use of the duration cue in their native language. As a rule, the L2 subjects demonstrated (1) category boundaries at longer durations, (2) larger width of category boundaries, and (3) lower consistency of responses compared to those of the L1 group. The perceptual strategies of L2 subjects might be based on the continuous auditory perception of the salient duration cue, or on the variable duration patterns associated with word stress in their L1, or on a combination of both strategies.

1. Introduction

Different languages exploit duration of speech sounds in different ways – in quantity languages duration has a function not available in non-quantity languages, namely to signal quantity contrasts. Typical examples of quantity languages are Finnish and Estonian, although Estonian has been claimed to undergo a prosodic change from quantity language to accent language (Lehiste, 2003). In Finnish there is a binary quantity contrast in both vowels and consonants, independent of each other and independent of stress (Suomi, Toivanen, & Ylitalo, 2008). The Finnish quantity contrast is best interpreted syntagmatically: contrastively long segments are interpreted as sequences of two identical phonemes, e.g. *kato* /kato/ 'dearth', *kaato* /kaato/ 'overturning', *katto* /katto/ 'roof' (Karlsson, 1969). Durational contrasts are even more important in Estonian where two types of quantity oppositions can be distinguished: (1) a binary opposition (short vs. long) at the phonemic level, and (2) a ternary opposition on the foot level, traditionally referred to as short (Q1), long (Q2), and overlong (Q3) quantity degrees. For example, Q1 *sada* /sata/ 'hundred', nom.sg.; Q2 *saada* /saata/ 'to send', sg.imperat.; Q3 *saada* /saa:ta/ 'to get'; Q1 *kala* /kala/ 'fish', nom.sg.; Q2 *kalla* /kalla/ 'arum', nom.sg.; Q3 *kalla* /kal:la/ 'pour', 2.sg.imperat. For a more detailed account, see Section 1.1.

Non-quantity languages do not use the duration cue contrastively; instead duration has somewhat different functions. For example, in Russian it acts as the main cue for word stress: vowels have a longer duration in stressed syllables than in unstressed ones (Bondarko, 1998). Word stress in Russian is not fixed; it can be on any syllable resulting in different lexical items.

When acquiring the phonological system of a language, native speakers of a quantity language learn to master durational variations in order to produce and perceive contrastive phonemic durational oppositions. As duration of speech segments in fluent speech varies to a great extent depending on a number of factors such as speech style, articulation rate, phonetic context and position of a segment in a word and in an utterance, morphological and syntactic as well as paralinguistic factors, the perception of the phonological length of a sound cannot be based on the absolute duration of a segment. Instead, a relative processing mechanism must be involved in perceptual mapping of speech segments into discrete phonological quantity categories, a mechanism that takes into account the plasticity of sound duration in variable contexts and the language-specific categorical representation of quantity.

Numerous studies have shown that the native language (L1) affects the acquisition of a second language (L2), thus one can

* Corresponding author. Tel.: +372 6204200; fax: +372 6204151.
E-mail address: einar@ioc.ee (E. Meister).

0095-4470/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2011.01.005

expect consequences in L2 speech perception and production due to the different roles of duration in different languages. Perception of phonological quantity by non-native speakers has been explored in several studies involving languages with different roles of duration, for example a study of Swedish quantity contrasts involved L1 speakers of Spanish, English, and Estonian (McAllister, Flege, & Piske, 2002), a study of Finnish involved L1 speakers of Russian (Ylinen, Shestakova, Alku, & Huotilainen, 2005), a study of Japanese involved English L1 subjects (Hayes-Harb, 2005), and a study on Estonian quantity perception involved native and non-native listeners (Lippus, Pajusalu, & Allik, 2009).

McAllister et al.'s (2002) experiments on the perception of Swedish quantity contrast by native speakers of languages with different role of duration (Estonian, English, and Spanish) showed that native Estonian subjects' performance in distinguishing Swedish quantity oppositions was very much like that of Swedish subjects, the native Spanish speakers were the least successful and the native English speakers' performance was somewhere between these two speaker groups. These findings suggest that learning Swedish quantity is related to the role of the duration cue in the L1, and are summarized in the Feature Hypothesis which states that "L2 features not used to signal phonological contrasts in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner's production of the contrast based on this feature" (McAllister et al., 2002, p. 230).

Ylinen et al.'s (2005) study was conducted to establish the short vs. long category boundary of Finnish vowels in stressed and unstressed positions. Testing the perception of native Finns and Russians showed that Finnish L1 speakers' categorization was more consistent and their short vs. long category boundary occurred at a significantly shorter vowel duration in comparison to the Russian subjects; in addition Finnish subjects showed a sensitivity peak in discrimination results occurring at the short/long category boundary, whereas Russian subjects showed continuous perception of duration and no evidence of a category-boundary effect. The results are in line with an earlier neurolinguistic study (Nenonen, Shestakova, Huotilainen, & Näätänen, 2003), which demonstrated significant differences in the mismatch negativity (MMN) amplitude for duration change in the case of speech and non-speech stimuli in native Finnish speakers, whereas in the same experiment Russian subjects showed no differences for duration change in speech and non-speech stimuli.

Hayes-Harb (2005) studied the identification of Japanese single vs. double (geminate) consonants by three groups of listeners: native speakers of Japanese, monolingual English speakers, and native English speakers learning Japanese. The results showed that native Japanese subjects exhibited categorical perception of consonant length, monolingual English subjects demonstrated continuous perception of consonant length, whereas English learners of Japanese exhibited results somewhere in between of the two preceding groups.

Lippus et al. (2009) studied the role of the pitch cue in the perception of Estonian quantity degrees by native and non-native listeners with different language backgrounds (Latvian, Finnish, and Russian). The results confirmed the importance of the pitch cue for native Estonian listeners in the perception of Estonian quantity degrees, especially in distinguishing long (Q2) and overlong (Q3) quantity degrees; if the pitch cue is not available, then in distinguishing the quantity degrees native speakers rely on the temporal structure only. Non-native listeners' perception was affected by the prosodic system of their first language. The Latvian prosodic system uses both temporal and tonal features; however, Latvian listeners tended to rely more on tone in distinguishing Q2 and Q3 than native Estonian listeners. Interestingly, in the case of missing tonal information, the Latvian listeners failed in exploiting the salient temporal cues in distinguishing Q2 and Q3. The Finnish listeners benefited from the short vs. long quantity oppositions of their L1 and categorized the Estonian quantity degrees by the temporal structure only; the pitch cue had no effect. Surprisingly, the Russian listeners showed results similar to those of the Finnish speakers, despite the fact that the Russian prosodic system does not exploit temporal nor pitch cues contrastively.

A different approach to L2 perception is based on a language independent, purely auditory processing model. The "desensitization" hypothesis proposed by Bohn (1995, p. 294) is based on results which "… suggest that duration cues in vowel perception are easy to access whether or not listeners have had specific linguistic experience with them". Bohn supposed that when subjects' L1 does not sensitize them to exploit spectral differences in distinguishing vowel contrasts, then durational differences will be used. The evidence for the hypothesis was found in a study involving the perception of American English vowels by native Spanish and German listeners, which demonstrated that L2 subjects tended to rely more on the duration cue than on spectral quality. The desensitization hypothesis has been supported by a number of subsequent studies involving native speakers of Russian and Spanish (Kondaurova & Francis, 2008), Catalan (Cebrian, 2006), etc. However, the results by McAllister et al. (2002), as already mentioned above, were consistent with the Feature Hypothesis and thus did not find support for the desensitization hypothesis.

The present study investigated the perception of Estonian duration-based phonological contrasts by non-native speakers with Russian-language background in comparison to native speakers of Estonian. In the experiments three stimulus sets involving short vs. long quantity oppositions in different phonetic contexts were used : (1) isolated vowels V vs. VV, (2) one-syllable nonsense words CVC vs. CVVC, and (3) disyllabic words CVCV vs. CVVCV. The first stimulus set was included in order to explore the question whether the short/long categorization in the case of isolated vowels is possible at all. In addition to investigating the short/long vowel contrast in CV(V)C context, the second stimulus set was designed also to study the role of speaking rate (manifested as variable duration of the initial consonant) on the short/long categorization. The third stimulus set addressed the role of pitch in the categorization of quantity contrasts in disyllabic words. Considering the distinct roles of duration in Estonian and Russian and the earlier findings discussed above, we hypothesized that the two subject groups would exploit different perceptual strategies in a short/long categorization task, resulting in differences in the following characteristics: (1) location of the category boundary, (2) width of the category boundary, and (3) consistency of categorization. The results are discussed in the context of the Feature hypothesis and the desensitization hypothesis.

1.1. Estonian quantity system

Estonian is known as a language with a three-way quantity system involving contrastive prosodic patterns traditionally referred to as short (Q1), long (Q2), and overlong (Q3) quantity degrees. The early treatments of Estonian quantity system (e.g. Ariste, 1939) postulated a three-way contrast based on segmental duration in the stressed syllable (i.e. all vowels and consonants occur in three contrastive quantities in this position). A large number of subsequent studies have shown that the system is much more sophisticated than this, and that it is certainly different from the quantity system of Dinka in which a three-way vowel length contrast has been reported (Remijsen & Gilley, 2008). Nowadays there is general agreement that the ternary phonological contrast in Estonian is characterized by a complex interaction of durational and tonal cues within a disyllabic foot (a sequence consisting of the stressed syllable and the following unstressed syllable) (e.g. Eek & Meister, 1997, 2003; Krull & Traunmüller, 2000; Lehiste, 1997, 2003). Quantity oppositions are manifested in vowels and diphthongs of the stressed syllable, e.g. Q1 *sada* /sata/ 'hundred', nom.sg.; Q2 *saada* /saata/ 'to send', sg.imperat.; Q3 *saada* /saa:ta/ 'to get'; Q2 *koera* /koera/ 'dog', gen.sg.; Q3 *koera* /koe:ra/ 'dog', part.sg., or in consonants and consonant clusters between the first and second syllable vowels, e.g. Q1 *kala* /kala/ 'fish', nom.sg.; Q2 *kalla* /kalla/ 'arum', nom.sg.; Q3 *kalla* /kal:la/ 'pour', 2.sg.imperat.; Q2 *lehma* /lehma/ 'cow', gen.sg.; Q3 *lehma* /leh:ma/ 'cow', part.sg. Although the duration of the vowel in the unstressed syllable varies to a great extent, there is no phonological length contrast in the unstressed syllable; vowels in unstressed syllables are classified as phonologically short. Due to the tendency towards foot isochrony the duration of the second syllable vowel is fully predictable, being shortest in Q3 and longest in Q1. There is no quantity opposition in word-initial consonants. In monosyllabic words the segmental-level duration oppositions may occur, e.g. *saag* /saak/ 'saw' nom.sg.; *saak* /saakk/ 'crop' nom.sg.; *sakk* /sak:k/ 'jag' nom.sg., however, all monosyllabic words are treated as being in the overlong quantity degree.

As the main feature best distinguishing the Estonian quantity oppositions, Lehiste (1960) has introduced the durational ratio first syllable/second syllable in the foot as follows: 2:3 for Q1, 3:2 for Q2, and 2:1 for Q3. Similar durational ratios have been confirmed in numerous later studies (Eek, 1974; Eek & Meister, 1997; Krull, 1991, 1992; Liiv, 1961); these ratios have been observed to remain stable even in spontaneous speech (Asu, Lippus, Teras, & Tuisk, 2009; Krull, 1993). In addition to durational ratios, the location of the fundamental frequency peak within the foot has been observed to be the second important cue contributing to the perception of the quantity degrees (Eek, 1980; Lehiste, 1960; Liiv, 1961). The typical pitch pattern for Q1 and Q2 is rising or flat during the first syllable and falling in the unstressed syllable, for Q3 the F0 peak is located around the first third of the stressed vowel and falls during the rest of the stressed syllable and in the unstressed syllable (Asu et al., 2009; Eek, 1980; Lehiste, 1960; Liiv, 1961).

The duration ratios of syllables work well as descriptors of the quantity degrees from the point of view of speech production but they fail in explaining the quantity oppositions from the perceptual point of view. Namely, the perception experiments by Lehiste and Fox (1992) have shown that native speakers of Estonian (and also native English speakers) are able to distinguish only two, not three different duration ratios, i.e. the duration ratios 3:2 (typical of Q2) and 2:1 (typical of Q3) are not distinguished.

Eek and Meister (1997, 2003) showed that native Estonian listeners were able to distinguish the stressed Q1 syllables from Q2 and Q3 syllables, but failed to discriminate between the stressed Q2 and Q3 syllables when these were presented to subjects without the second syllable, although the durations and F0 curve of the stressed syllables were kept intact. (In the experiment the second syllable of natural Q2 and Q3 disyllabic words was cut off.) The evidence thus suggests that, on the segmental (and syllabic) level only a short vs. long distinction can be made; the three-way distinction is possible only in disyllabic units. Based on these results, Eek and Meister (2003, 2004) proposed a model, which describes the production and perception of Estonian quantity oppositions by the durational ratios of adjacent segments; a similar approach has been proposed also by Traunmüller and Krull (2003).

A recent study by Kalvik and Mihkla (2010) investigated different characteristics possibly important for the realization of quantity contrasts in fluent speech and compared the two models (i.e. the classical model based on syllable duration ratios and the alternative model based on durational ratio of adjacent segments) using statistical methods. The classical model outperformed the alternative model by explaining more than 75% of the variability of the data in contrast to the 58% achieved by the alternative model.

In several languages vowel category contrasts involve both spectral and temporal cues (e.g. the tense-lax contrast in English). This is not the case for Estonian vowels in the contrastive quantity degrees. The spectral differences of Estonian stressed vowels in Q1, Q2, and Q3 feet do not exceed 1 Bark in read speech (Eek & Meister, 1998). In spontaneous speech the quality variations are more extensive—vowels in the stressed syllable of Q2 and Q3 feet are more peripheral and correspond roughly to those of read speech, whereas vowels in the stressed syllable of Q1 feet are much more centralized in comparison to those of read speech (Lippus, 2010). The quality of vowels in unstressed syllables tends to be reduced and varies significantly (unstressed vowels of Q3 feet being centralized most and the unstressed vowels of Q1 feet least (Eek & Meister, 1998)), but most of the variation is not connected with the quantity of the foot (Lippus, 2010).

In the current paper we focus only on the short vs. long opposition at the phonemic level, which is also a primary acoustic cue in distinguishing the short (Q1) quantity degree from the long (Q2), and overlong (Q3) quantity degrees; however, in this paper no attempt is made to exploit the results for the interpretation of the whole system of Estonian quantity degrees.

1.2. *Duration of Estonian short and long vowels: acoustic data*

Acoustic measurements of Estonian speech segments have been reported in a great number of studies (mainly in the context of word prosody), from the seminal works by Lehiste (1960), Liiv (1961), and Eek (1974) to more recent studies by Krull (1997), and Eek and Meister (1998, 2003), among others. The data from Eek and Meister (2003) is summarized in Table 1.

According to Liiv's data (1961), obtained from words embedded in sentences read in laboratory conditions by five subjects, the duration of short vowels in the stressed syllable of Q1 feet ranged from 109 to 131 ms, and that of long vowels from 178 to 225 ms when measured in Q2 feet, and from 219 to 260 ms when measured in Q3 feet. The duration of vowels in unstressed syllable varied from 137 to 180 ms when measured in Q1 feet, from 117 to 135 ms and from 81 to 102 ms when measured in Q2 and Q3 feet, respectively. The mean short/long stressed vowel duration ratio was 1:1.7 in the comparison of the stressed vowels of Q1 and Q2, and 1:2.0 in the comparison of the stressed vowels of Q1 and Q3.

Eek (1974) provided measurement data of simple word structures representing the quantity oppositions Q1, Q2, and Q3 in sentence-initial position read by one male speaker. The mean duration of short vowels in the stressed syllable of Q1 feet was 104 ms and that of long vowels 203 and 243 ms when measured

Table 1
Mean duration of vowels in one-syllable CVCC, CVVC, and CVVCC words and that of the stressed vowel in the two-syllable CVCV, CVVCV, and CVV:CV words at fast (F), moderate (M), and slow (S) articulation rate (Eek & Meister, 2003, Tables 6 and 9).

	Mean duration of V(V:) (in ms)					
Articulation rate	cVcc	cVVc	cVVcc	cVcv	cVVcv	cVV:cv
F	72	213	153	73	136	181
M	90	286	207	93	172	238
S	125	401	286	129	254	371

in Q2 and Q3 feet, respectively. The short/long ratio was 1:2.0 when comparing vowels of Q1 and Q2, and 1:2.3 when comparing vowels of Q1 and Q3. The mean duration of unstressed vowels was 141 ms for Q1, 100 ms for Q2, and 62 ms for Q3.

More recent data provided by Eek and Meister (1998) were measured from two-syllable words (CVCV, CVVCV, and CVV:CV) in sentences read by three male speakers. The mean duration of stressed short vowels in Q1 feet was in the range 69–97 ms, that of long vowels in the ranges 141–182 and 182–251 ms when measured in Q2 and Q3 feet, respectively. The mean short/long ratio was 1:2.0 when comparing the stressed vowels of Q1 and Q2, and 1:2.6 when comparing the stressed vowels of Q1 and Q3. The mean duration of unstressed vowels was in the ranges 70–121, 76–93, and 67–85 ms when measured in Q1, Q2, and Q3 feet, respectively.

Another study by Eek and Meister (2003) reported vowel durations measured in one-syllable CVCC, CVVC, and CVVCC Q3 words as well as in two-syllable CVCV, CVVCV, and CVV:CV words representing quantity degrees Q1, Q2 and Q3, respectively. The studied words were embedded in different positions of the carrier sentences resulting in four different prosodic contexts (sentence-initial, sentence-final, focus, non-focus). The sentences were read in the form of dialog at slow, moderate, and fast articulation rates by two male and two female native Estonian speakers. The mean duration of the short stressed vowels in the words studied (Table 1) varied in the range 72–129 ms and that of the long vowels in the range 136–401 ms, depending on word structure and articulation rate.

Despite the large variation in vowel duration due to different articulation rates, the short/long duration ratio remained almost constant and independent of articulation rate, ranging from 1:3.0 to 1:3.2 when comparing word structures CVCC and CVVC, from 1:2.1 to 1:2.3 when comparing word structures CVCC and CVVCC, from 1:1.8 to 1:2.0 when comparing word structures CVCV and CVVCV, and from 1:2.5 to 1:2.9 when comparing word structures CVCC and CVV:CV.

Krull (1997) investigated temporal characteristics of Estonian quantity degrees in the conversational speech of three native speakers, comparing words in prepausal and non-prepausal positions. The mean short/long vowel duration ratio in conversational speech was somewhat less than that in read speech as discussed above, ranging from 1:1.6 for prepausal to 1:1.8 for non-prepausal position when comparing stressed vowels of Q1 and Q2, and from 1:2.5 for prepausal to 1:2.3 for non-prepausal position when comparing stressed vowels of Q1 and Q3. The mean duration of stressed vowels in Q1 spanned the range 71–77 ms depending on the position of the word, and the duration of stressed vowels in Q2 and Q3 was 124–126 and 162–190 ms, respectively. The duration of unstressed vowels ranged from 63 to 99 ms in non-prepausal and from 96 to 121 ms in prepausal position; the longer vowel duration in prepausal position being caused by final lengthening.

In the most recent study of the temporal characteristics of spontaneous Estonian (Asu et al., 2009), the duration of the stressed short vowel in Q1 feet was in the range 60–70 ms, that of the long stressed vowels 130–160 ms in Q2 feet, and 140–170 ms in Q3 feet. The corresponding duration ratios of short and long vowels were 1:2.2 and 1:2.4 when comparing stressed vowels of Q1 and Q2, and Q1 and Q3, respectively.

To summarize the findings of the studies reviewed above, Estonian phonologically long vowels are 1.6–2.0 times longer than phonologically short vowels when comparing the stressed vowels of Q2 and Q1, and 2.1–2.9 times longer when comparing the stressed vowels of Q3 and Q1. In the stressed syllable the absolute duration of a phonologically short vowel ranges from 60 to 129 ms and the duration of phonologically long vowel spans from 124 to 371 ms; in an unstressed syllable the duration of a vowel varies from 62 to 227 ms (as already mentioned above, vowels in unstressed syllables do not produce quantity oppositions and they are considered phonologically short).

Similar results have been reported for other quantity languages, like Finnish and Japanese. Pooling durational data from different studies of Finnish (Suomi, 2005, 2006, 2007, 2009; Suomi & Ylitalo, 2004; Suomi, Toivanen, & Ylitalo, 2003) shows that the duration of phonologically short stressed vowels ranges from 56 ms in unaccented words to 87 ms in contrastively accented words, and that of phonologically long vowels from 129 to 212 ms under corresponding accentual conditions; the computed mean short/long ratio is 1:2.4. A study of Japanese vowel length reports the mean short/long vowel duration ratio 1:2.5 for stressed and 1:2.2 for unstressed vowels pooled across fast, normal and slow speaking rates (Hirata, 2004).

1.3. Perception data: short/long category boundary

The acoustic data reviewed above show that in quantity languages short and long vowel categories are well distinguished in speech production—the durational ratio of short and long vowels around 1:2 is quite stable even at different speaking rates and should guarantee reliable perceptual distinction of these phonological categories.

Eek and Meister (2003) studied the perceptual boundary of short and long vowel categories of native Estonian subjects in one-syllable synthetic CV(V)C nonsense words (*sas* vs. *saas*) by varying the duration of the vowel (from 40 to 250 ms in 10 ms steps) at three speaking rates, where the duration of the word-initial consonant (70, 100, and 130 ms) was assumed to play the role of a cue for speaking rate. It was observed that the short/long category boundary is a function of speaking rate—the mean category boundary occurred at a vowel duration of 100 ms in the case of C1 = 70 ms, at 120 ms in the case of C1 = 100 ms, and at 140 ms for C1 = 130 ms.

A recent study on the perception of quantity degrees in Estonian using re-synthesized disyllabic CV(V)CV words with manipulated duration of the vowel in the first syllable (Lippus and Pajusalu, 2009) yielded the short/long boundary at 115–130 ms vowel duration depending on the regional dialectal background of the subjects.

In Finnish, the short/long category boundary has been studied in CV(V)CV(V) words by manipulating the duration of vowels in the first and second syllable (in contrast to Estonian, in Finnish short/long oppositions occur in both stressed and unstressed syllables) resulting in two continua: (1) from /tu:ku/ to /tuku/ and (2) from /tuku:/ to /tuku/ (Ylinen et al., 2005); assuming the standard syntagmatic phonological interpretation of Finnish quantity, these oppositions would be represented as /tuuku/ and /tuku/, /tukuu/ and /tuku/. In the experiment four groups of subjects including native speakers of Finnish, Russian L2 users of Finnish with short and long exposure, and naïve Russians with no exposure to Finnish acted as participants. For native Finns the perceptual short/long category boundary was at 116 ms in the first-syllable position and at 162 ms in the second-syllable position. In all L2 groups the boundary occurred at a somewhat longer duration (+7–17 ms) in both syllable positions, in such a way that for the L2 group with long exposure the boundary was closest to the L1 speakers; the native Russians and the L2 group with short exposure exhibited almost identical results.

1.4. Vowel duration in Russian

In Russian, duration has no phonological function; instead it serves as the main cue for word stress. The duration of stressed-syllable vowels is longer than the duration of vowels in unstressed syllables; the degree of reduction of unstressed vowels increases with the distance of a vowel from the location of stress, the durational reduction of a vowel is also accompanied by qualitative reductions, especially in the case of the vowel /a/ (Bondarko, 1998).

Unlike Estonian, where, as a rule, the main stress is associated with the initial syllable of the word (except for some loanwords), Russian is a language with free (or moving) word stress. Stress in Russian is lexically determined, and there are minimal pairs like /ˈzamok/ 'castle' vs. /zaˈmok/ 'lock', /ˈmuka/ 'suffering' vs. /muˈka/ 'flour', etc.

Comparing the perception of word stress by Estonian and Russian subjects, Eek (1987) concluded that stress perception by Estonian subjects is mainly associated with the F0 pattern whereas for Russian subjects duration serves as the main cue. Consequently, the different roles of duration in Estonian and Russian may confuse Russian subjects as they may perceive Estonian phonetically long unstressed vowels as stressed (similarly to Finnish, as observed by De Silva (1999)).

Temporal characteristics of speech segments in spontaneous and read speech in Russian have been reported in a number of papers (Bolotova, 2003; Bondarko, Volskaya, Tananaiko, & Vasilieva, 2003; De Silva, Iivonen, Bondarko, & Pols, 2003; Skrelin, 2004; Volskaya & Skrelin, 2004). Pooling across the studies just mentioned, the mean duration of Russian stressed vowels is reported to be 84 ms (SD=38 ms) in spontaneous and 96 ms (SD=40 ms) in read speech; for unstressed vowels the corresponding figures are 59 ms (SD=35 ms) and 64 ms (SD=30 ms). In spontaneous speech the stressed vowels in the intonation center (focus, prominence) are on average 24 ms longer than stressed vowels outside of the focus, in read speech this difference is even greater, ranging up to 39 ms longer vowel duration in pitch-prominent position (Bolotova, 2003). The variability of vowel durations (due to variable speech rate) has been found to be greater in spontaneous speech than in read speech. Faster local tempo in spontaneous speech results in the increase of vowel elision in both stressed and unstressed positions and in vowel duration alterations in stressed position, whereas in read speech vowel durations tend to alter more in unstressed position (Skrelin, 2004).

2. Experiments

2.1. Subjects

Two groups of subjects were recruited: native speakers of Estonian (referred to below as the L1 group) and non-native speakers with Russian as their first language (referred to as the L2 group); both groups consisted of 10 speakers (5 male and 5 female). The L1 subjects (age 21–54, median 26.5) came from monolingual Estonian families living in the capital area; most of them were also born in the same area (2 subjects were born in the southern part of Estonia and continued to live there till adulthood). All L1 subjects had acquired or are currently acquiring a university degree; they represent standard Estonian pronunciation and have no reported hearing problems.

The L2 subjects (age 21–34, median 24.5) were born in monolingual Russian families living mostly in north-east Estonia or in the capital area (one subject (Sp 9) was born in Russia, and her family moved to Estonia when she was 16); they were educated in Russian-language basic and high schools and have acquired or are currently acquiring a university degree in Estonia (one subject (Sp 1) acquired his university degree in St. Petersburg). Most of the L2 subjects started to learn Estonian at school at the age of 5–13, one subject (Sp 9) at the age of 16 and one (Sp 1) at the age of 20. All L2 subjects use Estonian almost every day at the university or at their place of work as they all have friends or colleagues who are native speakers of Estonian, but at home they communicate in Russian (except one subject (Sp 1)). In a self-assessment of the proficiency of Estonian (on a scale with options "basic", "average", "good", and "very good") two subjects rated themselves as "very good", five subjects as "good", and three subjects as "average". None of the subjects reported any language impairment (see Table 2 for summary of L2 subjects).

Both subject groups first participated in the perception experiments and then were recorded while reading an Estonian text corpus for further studies. All subjects were paid for their participation in the study.

2.2. Stimulus sets

The identification of short vs. long quantity categories was studied by using three different stimulus sets. All stimuli were created from the base units /aa/, /saas/ and /saada/ (in the Q2 quantity degree) pronounced by a native Estonian male speaker. The duration and F0 manipulations of the stimuli were performed by using the PSOLA-facility of the Praat software (Boersma & Weenink, 2009).

The first set was an isolated vowel set (V-set) including instances of the Estonian vowel /a(a)/ with variable duration. In reality, there are no minimal pairs involving single phoneme length in Estonian—a short isolated /a/ could be perceived as a nonsense vocalic segment cut out from a larger stretch of speech, but the long isolated /aa/ represents a monosyllabic word Aa /aa:/, a place name in North-East Estonia, nom.sg. Thus, the short vs. long opposition in the V-set also constitutes a contrast between an impossible word and a real word. The V-set was included in order to see whether, despite this latter contrast, listeners are able to identify the short vs. long opposition in isolated vowels.

The stimulus continuum of the V-set was created by manipulating the duration of the base vowel from 90 to 200 ms in 10 ms steps. The fundamental frequency of the stimuli was set to 130 Hz at the beginning, after which it descended linearly towards the final target of 120 Hz. The number of stimuli in the V-set was 12.

The second set was the CVC-set including a continuum of nonsense words /sa(a)s/ with variable vowel duration resulting in a contrast between /sas/ and /saas/. Using the opposition CVC vs. CVVC might look doubtful as the short vs. long vowel contrast in monosyllabic words is always combined with a reversed opposition of the following consonant; i.e. the phonologically possible structures are CVCC and CVVC, but not CVC. Consequently, the short vs. long opposition in the CVC-set additionally involves a

Table 2
Characteristics of L2 subjects (Age=chronological age, in years; AOL=age when subjects started to learn Estonian, in years; L2 use=self-reported use of Estonian; L2 proficiency=self-reported proficiency of Estonian).

Speaker	Gender	Age	AOL	L2 use	L2 proficiency
Sp 1	M	34	20	Daily	Good
Sp 2	M	28	12	Daily	Average
Sp 3	M	21	6	Daily	Average
Sp 4	M	25	6	Daily	Good
Sp 5	M	21	7	Daily	Very good
Sp 6	F	20	5	Daily	Good
Sp 7	F	33	11	Daily	Average
Sp 8	F	24	13	Frequently	Good
Sp 9	F	27	16	Daily	Very good
Sp 10	F	24	8	Daily	Good

contrast between a non-word and a possible word. However, as previous experiments have shown (Eek & Meister, 2003), this latter contrast does not interfere with the perception of vowel length in the structures involved.

The duration of the vowel in the CVC-set ranged from 80 to 160 ms in 10 ms steps, the F0-contour modeled that of a natural pronunciation starting at 130 Hz, rising to 135 Hz during the first third of the vowel duration and then lowering to the final target of 120 Hz during the rest of the vowel. The duration of initial and final consonants was set to 110 ms in the first subset, referred as CVC(1), and to 80 ms in the second subset, referred as CVC(2), in order to model two different speech rates (as shown in Eek and Meister (2003), the duration of word-initial consonant acts as a cue for local speaking rate). In the case of the shorter consonant durations we expected the location of the short/long boundary to occur at a shorter vowel duration. The total number of stimuli in the CVC-set was 18.

The third stimulus set was the CVCV-set consisting of stimuli with varying duration of the first-syllable vowel, resulting in a continuum from the lexical item *sada* /sata/ 'hundred, nom.sg' in quantity degree Q1 to the lexical item *saada* /saata/ 'to send, sg.imperat.' in quantity degree Q2. The duration of V1 was manipulated from 80 to 160 ms in 20 ms steps, C1 and C2 were set to the constant value of 80 ms, and the duration of V2 was fixed to 140 ms. These durations of the stimulus segments are based on the acoustic measurements of the CV(V)CV structures in the Tempo-corpus (Eek & Meister, 2003), corresponding to moderate or slightly slower articulation rate. Two subsets with a different F0-contour were created in order to compare the role of pitch in the perception of quantity patterns in L1 and L2 listener groups. In the first subset, referred to as CVCV(1), F0 was kept constant at 130 Hz in both V1 and V2. In the second subset, referred to as CVCV(2), F0 was modeled to follow its typical course in natural Q1 and Q2 words, i.e. rising from 120 to 140 Hz during the first vowel and falling from 120 to 100 Hz during the second vowel. The number of different stimuli in the CVCV-set was 10.

2.3. Testing procedure

All perception tests were conducted in a quiet room and stimuli were presented to subjects via high-quality headphones from a laptop computer. The test was administered with Praat's (Boersma & Weenink, 2009) multiple forced-choice test facility. The test was organized in three blocks. First the isolated vowel stimulus set was presented, followed by the CVC and CVCV stimulus sets, with optional short breaks between blocks. Each stimulus of the first and second set was repeated five times and the stimuli of the CVCV-set were repeated three times in random order; balanced permutation for stimulus randomization was applied. In all tests listeners had to decide on vowel quantity in a binary identification task and answer by clicking in one of two response boxes on the screen. The boxes were labeled as "V" and "VV", as "CVC" and "CVVC", and as "CVCV" and "CVVCV" in three test blocks. A total of 180 stimuli were presented to the subjects. The duration of the test was approximately 15–20 min.

2.4. Data analysis

The results of each test block were presented as plots of the proportion of "short" responses against stimulus duration. Individual categorization functions were obtained by interpolating the response data of each subject with a probit function and the locations of individual short/long category boundaries (expressed in milliseconds) were calculated as the 50% cross-over point of the fitted curve. The average categorization functions and the average short/long category boundaries for L1 and L2 groups were established in a similar way. In addition, the durations corresponding to the values 0.8 and 0.2 of the fitted categorization curve were determined and the width of category boundary was calculated as the difference of these durations. The location and width of individual category boundaries in all stimulus sets (as dependent variables) were subjected to a quantitative analysis using one-way ANOVAs with language background or stimulus set as independent variables. The Tukey Honest Significant Difference test was used as a post-hoc test. To evaluate the reliability of raters, the intraclass correlation measure ICC1[1] (Shrout & Fleiss, 1979), which describes the consistency of responses within a group was calculated.

3. Results

3.1. V-set

The short/long categorization task involving isolated vowels was rather artificial since in Estonian there is no true short/long opposition in isolated vowels. The task was certainly complicated because there was no context available which would give the reference of local speech rate, and subjects should rely on their internal criteria only. Nevertheless, the categorization strategies of different subject groups might be different due to different roles of duration in their native language. Estonian subjects may exploit their language-specific tuning of duration processing (as has been suggested by Nenonen et al. (2003) for native speakers of Finnish), and Russian subjects in turn might use a general auditory processing mechanism of an acoustic continuum (Kuhl & Miller, 1978) and categorize stimuli on the basis of absolute duration. Due to latent categorization mechanisms the results of the task might demonstrate variable consistency in subjects' responses.

Despite the complicated task, most subjects were able to perform the categorization. The average categorization curves for V-set (Fig. 1) show that the L1 and L2 subjects behaved differently. All L1 subjects demonstrated consistent categorical perception: ICC = 0.94 (Fig. 2).

In the L2 group one subject (Sp 4) perceived most stimuli as "short", perceiving only one of the five repetitions of the two longest stimuli as "long". As this subject's responses did not show any categorization effect, his results were excluded from further analysis. The rest of the L2 subjects showed good categorization results with a consistency close to the L1 subjects: ICC = 0.93.

In the L1 subjects, the individual locations of the short/long category boundary derived from the probit-models varied from 131.3 to 148.7 ms, with 139 ms (SD = 6.2 ms) as the mean category boundary. The width of the category boundary ranged from 11.2 to 24.1 ms, with a mean of 14.8 ms (SD = 4.9 ms).

The locations of individual category boundaries of L2 subjects varied from 135 to 155 ms, and the mean category boundary was 146 ms (SD = 6.8 ms). The width of the category boundary for this group was much larger than that of the L1 group, ranging as it did from 9.3 to 34.5 ms, the mean being 21.2 ms (SD = 7.4 ms).

A one-way ANOVA confirmed the group effect for the category boundary [$F(1,17) = 5.7977$, $p < 0.05$], an ANOVA for

[1] The intraclass correlation coefficient is calculated using the ICC-function of the psych-library of R. It implements the methods by Shrout and Fleiss (1979) providing 6 different ICC-measures; in our paper we use ICC1 in which each target is rated by a different judge and the judges are selected at random. ICC1 is sensitive to differences in means between raters and is a measure of absolute agreement.

width of the category boundary also showed a significant effect [$F(1,17)=4.9308$, $p<0.05$].

3.2. CVC-set

The categorization task in the CVC-stimuli was more straightforward and realistic since the information about local speech rate was given by the consonantal context. The stimuli of our CVC vs. CVVC paradigm – one-syllable words /sas/ vs. /saas/ – were nonsense words in both Estonian and Russian, which provided equal test conditions for both subject groups. Moreover, in neither language is the CVC vs. CVVC opposition possible because in Estonian CVC words do not occur (all monosyllabic words are in Q3).

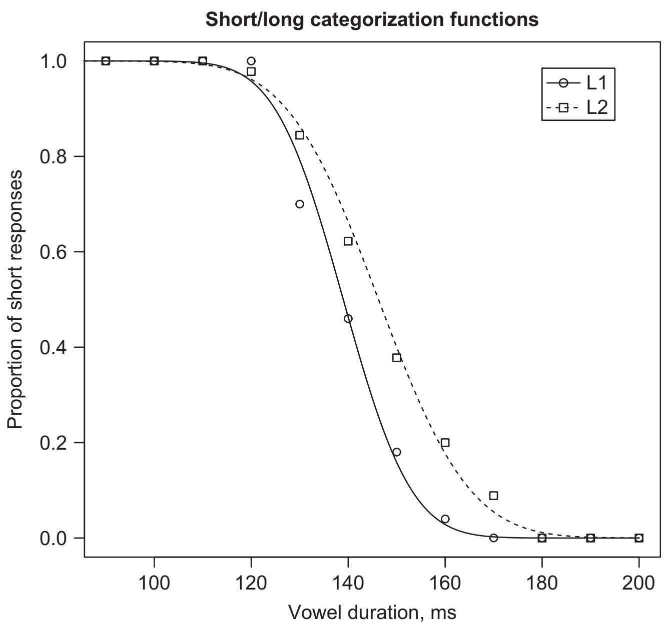

Fig. 1. Averaged categorization functions of L1 (solid line) and L2 (dotted line) groups in the V-set. Symbols ("○" for L1 and "□" for L2) represent the mean score of "short" responses to each stimulus; lines represent the fitted curves based on the logistic regression (probit) models.

In Estonian, the duration of a word-initial consonant acts as a cue for local speaking rate (Eek & Meister, 2003), and thus it has an effect on the short/long categorization of the following vowel—when the duration of the word-initial consonant is longer the location of short/long boundary is expected to occur at longer vowel duration. In similar experiments (Eek & Meister, 2003) with Estonian subjects only, the short/long category boundary of a vowel in the CVC-context occurred at a vowel duration 1.1–1.4 times longer than the duration of the word-initial consonant.

In the perception test involving the two CVC-sets (CVC(1) with C1=C2=110 ms and CVC(2) with C1=C2=80 ms), modeling slow and moderate speaking rate, all L1 and L2 subjects showed a consistent categorization effect (L1 group: ICC=0.95 for CVC1 and ICC=0.91 for CVC2; L2 group: ICC=0.9 for CVC1 and ICC=0.91 for CVC2).

In comparison to the V-set, in both CVC-sets the short/long category boundary occurred at much shorter vowel durations. In the CVC(1)-set, the individual short/long category boundaries for the L1 group varied from 109.5 to 123.1 ms, the mean category boundary being 117.5 ms (SD=4.8 ms). In the CVC(2)-set, for the same L1 group, the individual category boundaries varied from 101.3 to 120.9 ms, the mean category boundary being 111.6 ms (SD=5.9 ms).

In the L2 group the individual category boundaries occurred at a longer vowel duration, namely 115–130.7 ms in CVC(1), the mean category boundary being 124.2 ms (SD=6.9 ms). As with the L1 subjects, the duration of surrounding consonants also affected the perception of short/long category boundary in the L2 subjects – in CVC(2) the individual boundaries varied from 108.8 to 126.9 ms and the mean category boundary occurred at 119.2 ms (SD=5.5 ms) (Fig. 4).

A one-way ANOVA showed that subject group had a significant effect on category boundary location in both CVC-sets [(CVC(1): $F(1,18)=6.4531$, $p<0.05$; CVC(2): $F(1,18)=8.67$, $p<0.01$)].

Due to different locations of category boundaries the categorization functions (Fig. 3) are displaced along the time axis, but

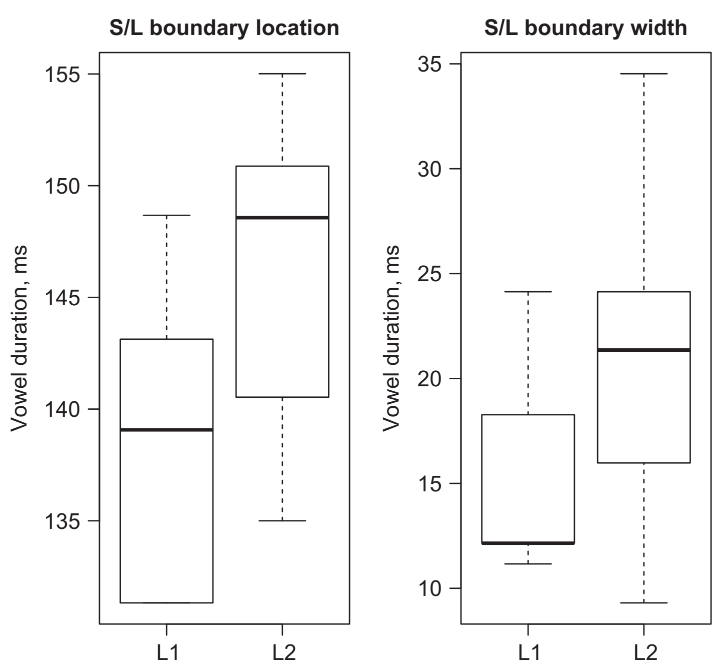

Fig. 2. Box-plots of short/long category boundary locations (left) and widths (right) in L1 and L2 subject groups in the case of V-set.

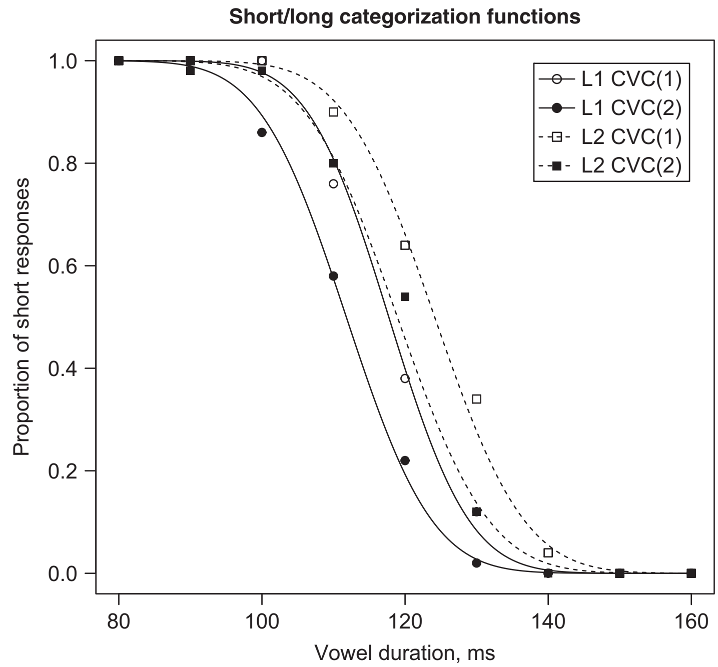

Fig. 3. Averaged categorization functions of L1 (solid lines) and L2 (dotted lines) groups in the stimulus sets CVC(1) and CVC(2). Symbols ("○" for L1 CVC(1), "●" for L1 CVC(2), "□" for L2 CVC(1), and "■" for L2 CVC(2)) represent the mean proportion of short responses to each stimulus, lines represent the fitted curves based on the logistic regression (probit) models.

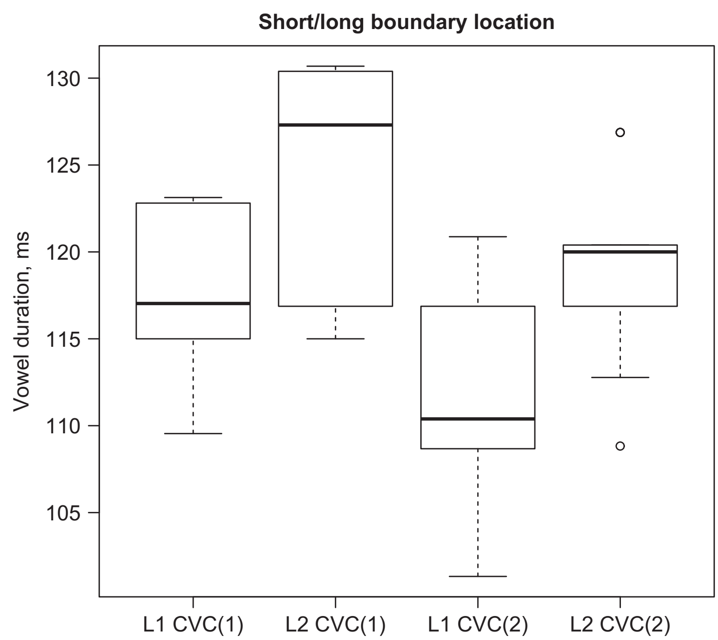

Fig. 4. Box-plot of short/long category boundary locations of L1 and L2 subject groups in CVC(1) and CVC(2) stimulus sets.

otherwise almost parallel. This predicts similar width values for both language groups in both CVC-sets. Indeed, the width of category boundary of L1 subjects is almost identical in both stimuli sets spanning from 2.3 to 18.3 ms in the CVC(1)-set and from 2.3 to 18.9 ms in CVC(2)-set, the mean values being 12.1 ms (SD=4.7 ms) and 11.4 ms (SD=5.2 ms), respectively. The widths of category boundary of the L2 subjects coincide well with those of the L1 subjects, with mean values of 10.5 ms (SD=6.6 ms) and 10.6 ms (SD=8.3 ms), respectively (Fig. 5). As expected, an ANOVA revealed no subject group effect on category boundary width in the CVC(1)-set [$F(1,18)=0.4245$, $p=0.5$] nor in the CVC(2)-set [$F(1,18)=0.0742$, $p=0.8$].

3.3. CVCV-set

The CVCV stimulus sets involved a continuum from the lexical item in quantity Q1 *sada* /sata/ 'hundred', nom.sg. to the lexical item in quantity Q2 *saada* /saata/ 'to send', sg.imperat. produced by changing the duration of the first vowel. The perception of Estonian quantity degrees is not based on the short/long opposition of the stressed-syllable vowel only; it depends on a more complex set of parameters including different durational and tonal patterns, picked up by native speakers in the process of language acquisition. For L2 speakers of Estonian the three-way quantity system often constitutes the hardest part of Estonian phonology.

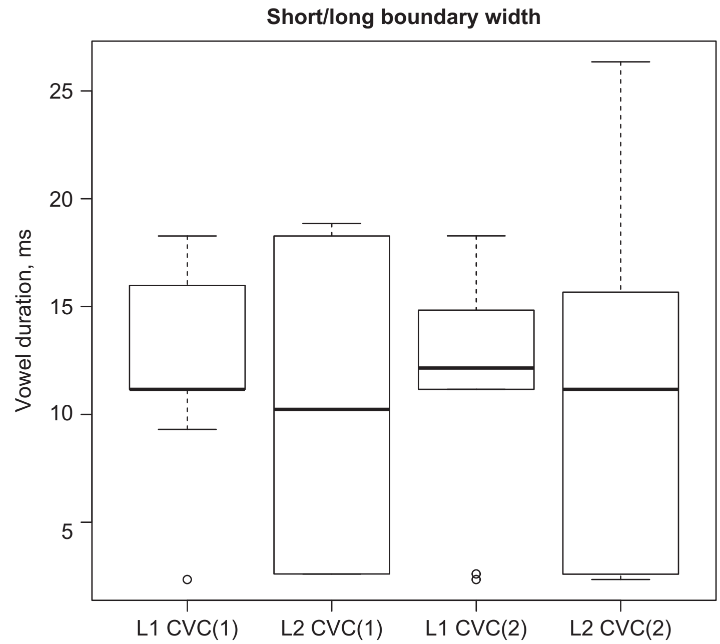

Fig. 5. Box-plot of short/long category boundary widths of L1 and L2 subject groups in CVC(1) and CVC(2) stimulus sets.

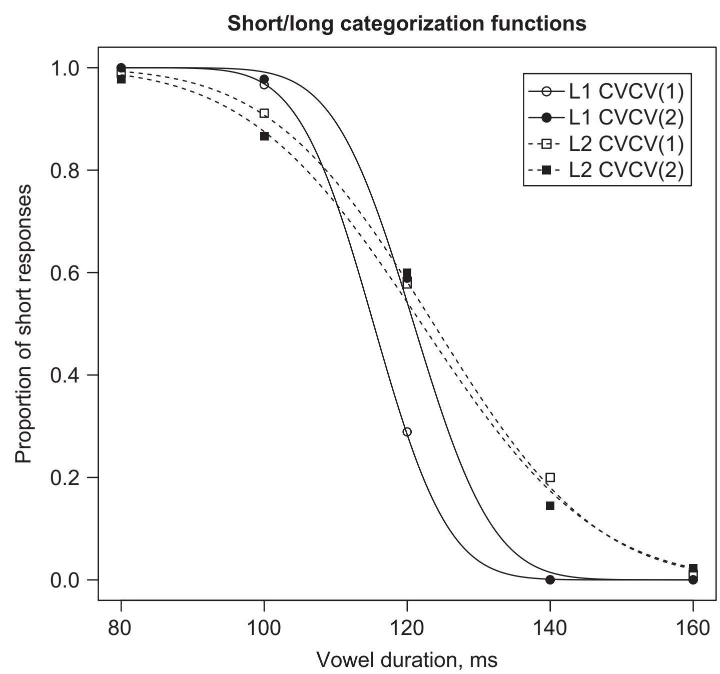

Fig. 6. Averaged categorization functions of L1 and L2 groups in the stimulus sets CVCV(1) and CVCV(2). Symbols ("○", "●", "□", and "■") represent the mean proportion of short responses to each stimulus, lines represent the fitted curves based on the logistic regression (probit) models.

In the two CVCV stimulus sets the primary contrast was based on the duration of the vowel in the first syllable, the duration of the other sounds was kept constant in all stimuli. In the CVCV(1)-set the complementary tonal feature was added by modeling a typical F0-contour of Q1 and Q2 words; in the CVCV(2)-set F0 was monotonous. In a short/long categorization task, the L1 subjects should benefit from the complementary tonal feature available in the CVCV(1)-set, and they were consequently expected to have the short/long category boundary location at a shorter vowel duration than in the CVCV(2)-set, where the categorization should be based on duration only. The L2 subjects were expected to rely, in both stimulus sets, on the duration cue only (as concluded in Lippus et al., 2009), and their short/long category boundaries should demonstrate no significant differences among the stimulus sets.

The results of the perception tests confirmed these hypotheses. As shown in Fig. 6, the average categorization functions of L1 and L2 groups look very different—the L1 curves are much steeper than the L2 curves, indicating a much narrower width of category boundary in the L1 group; the L1 curves for different stimulus sets are displaced in the time domain, suggesting a large difference between the category boundary locations in two stimulus sets, in contrast to the L2 curves (Fig. 7).

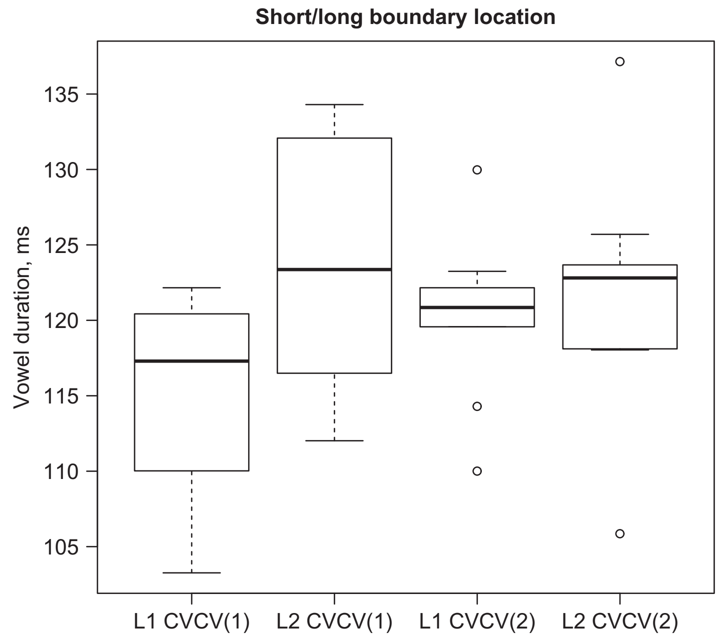

Fig. 7. Box-plot of short/long category boundary locations of L1 and L2 subject groups in CVCV(1) and CVCV(2) stimulus sets.

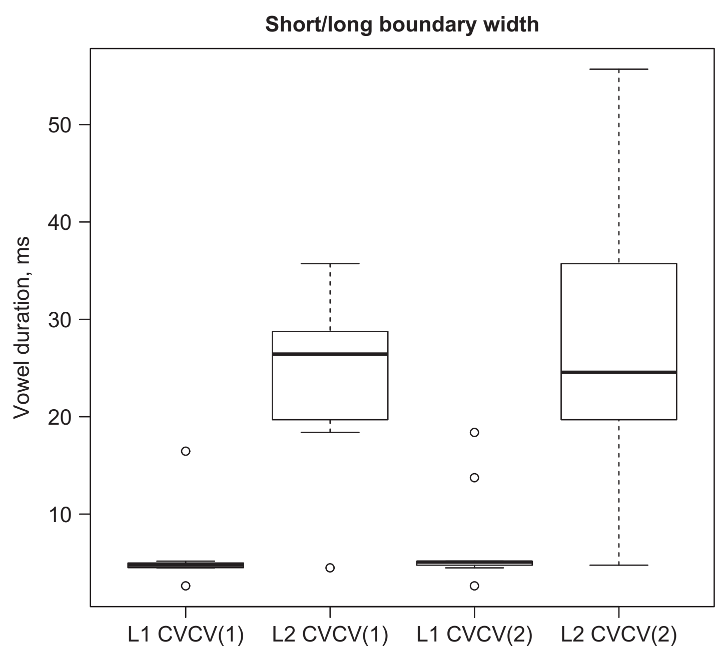

Fig. 8. Box-plot of short/long category boundary widths of L1 and L2 subject groups in CVCV(1) and CVCV(2) stimulus sets.

In the CVCV(1)-set the locations of individual short/long category boundaries in the L1 group varied from 103.2 to 122.2 ms, the mean category boundary being 114.6 ms (SD=7.1 ms). In the CVCV(2)-set the individual category boundaries varied from 110 to 130 ms, the mean being 120.2 ms (SD=5.3 ms). The variation of boundary locations of L2 subjects extended from 116.5 to 134.3 ms in CVCV(1) and from 105.9 to 137.2 ms in CVCV(2), with the means 124.1 ms (SD=8.9 ms) and 121.9 ms (SD=7.8 ms), respectively.

A one-way ANOVA on the location of the category boundaries of L1 and L2 groups showed a significant effect only in the CVCV(1)-set [$F(1,18)=6.9998$, $p<0.05$]; analysis of variance within the subject groups showed a marginally significant effect of stimulus set only in the L1 group [$F(1,18)=3.9983$, $p=0.06$]. However, a post-hoc test showed that there is a significant difference in the location of category boundaries between the L1 and L2 groups in the case of CVCV(1)-set ($p<0.05$), but not between the stimuli sets within the L1 group ($p=0.34$).

The width of category boundary appeared very different for the L1 and L2 groups (Fig. 8). In the L1 subjects the width varied from 2.6 to 16.5 ms in the CVCV(1)-set and from 2.6 to 18.4 ms in the CVCV(2)-set, with mean widths of 5.7 ms (SD=3.8 ms) and 6.9 ms (SD=5 ms), respectively. In the L2 subjects the corresponding measures were much larger, spanning from 4.5 to 35.7 ms in CVCV(1) and from 4.8 to 55.7 ms in CVCV(2), the mean widths being 24.3 ms (SD=8.8 ms) and 27 ms (SD=14.1 ms).

An ANOVA of category boundary width revealed a strong effect of subject group in both stimulus sets [CVCV(1): $F(1,18)=37.741$,

Table 3
The location and width of short/long category boundary (in ms) of native (L1) and non-native (L2) speakers of Estonian in all stimulus sets.

Short/long category boundary: location and width (in ms)										
Category boundary	V-set		CVC(1)-set		CVC(2)-set		CVCV(1)-set		CVCV(2)-set	
	L1	L2	L1	L2	L1	L2	L1	L2	L1	L2
Location	138.9	146.0	117.5	124.2	111.6	119.2	114.6	124.1	120.2	121.9
SD of location	6.2	6.8	4.8	6.9	5.9	5.5	7.1	8.9	5.3	7.8
Width	14.8	21.2	12.1	10.5	11.4	10.6	5.7	24.3	6.9	27.0
SD of width	4.9	7.4	4.7	6.6	5.2	8.3	3.8	8.8	5.0	14.1

$p<0.001$; CVCV(2): $F(1,18)=18.04$, $p<0.001$], but no effect of stimulus set within the subject groups.

The L1 subjects showed more consistent categorization and higher inter-speaker correlation in both stimuli sets than the L2 subjects (L1 group: ICC=0.9 for CVCV1 and ICC=0.89 for CVCV2; L2 group: ICC=0.8 for CVCV1 and ICC=0.81 for CVCV2).

4. Summary and general discussion

The main goal of the study was to explore the perception of Estonian short vs. long quantity oppositions by native Estonians and non-native subjects with Russian as their L1. The perception of Estonian quantity distinctions was examined using vowel continua from the short (V) to the long (VV) category in three contexts: (1) V-set, including isolated vowels V vs. VV, (2) CVC-set, involving one-syllable nonsense words CVC vs. CVVC, and (3) CVCV-set, involving two-syllable words CVCV vs. CVVCV. The CVC-set had two subsets with different duration of surrounding consonants; the CVCV-set involved two subsets with natural and monotonous F0 curve, respectively.

The results demonstrated differences in short/long category perception by L1 and L2 subject groups in all three stimulus sets. The category boundaries of L1 subjects always occurred at shorter durations compared to the boundary locations of L2 subjects. The widths of category boundaries of the L1 group were significantly narrower in the V-set and the CVCV-sets in comparison to those of L2 group, whereas in CVC-sets the widths were almost identical for both subject groups. As a rule, in all stimulus sets, the consistency of responses by the L2 group was slightly lower than the consistency of the L1 group. All test results are summarized numerically in Table 3 and graphically in Fig. 9.

Our findings are very much in line with the studies on perception of phonological quantity in Finnish by native speakers of Finnish, naïve Russians and L2 speakers of Finnish with Russian as their L1 using CV(V)CV(V) and isolated vowel continua (Ylinen et al., 2005; Ylinen, Shestakova, Huotilainen, Alku, & Näätänen, 2006). Thus the category boundaries, the steepness of categorization functions (in our study, however, instead of the slope of categorization curve, the width of category boundary was used) and the consistency of responses of L2 subjects all exhibited similar trends in both Finnish and Estonian. The similarity of the results was highly expected as the short/long category distinction is, in both languages, based on the contrastive use of segment duration with the main difference that in Finnish short/long category contrasts do occur in both stressed and unstressed positions whereas in Estonian the contrasts are possible only in stressed position.

In the V-set the category boundary of both L1 and L2 subject groups was located at much longer durations than in the CVC- and CVCV-sets. This can be explained by the missing phonetic context that would provide a reference of local speech rate which is necessary in judging the durational characteristics of speech

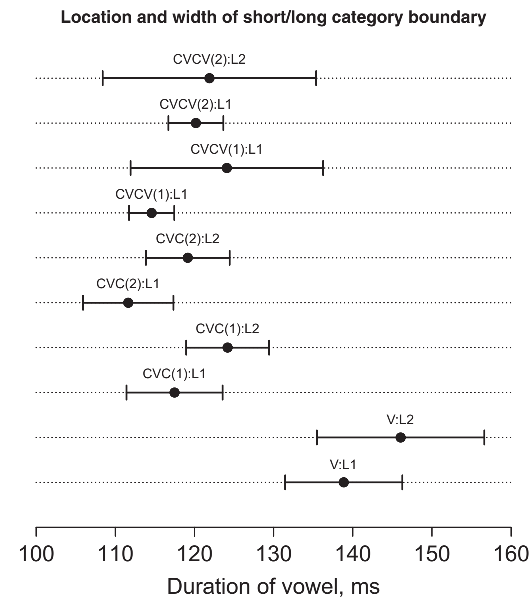

Fig. 9. The location ("●") and width (shown as whiskers) of short/long category boundary of native (L1) and non-native (L2) speakers of Estonian in all stimulus sets.

sounds (on the role of phonetic context in speech perception, see e.g. Kingston, Kawahara, Chambless, Mash, & Brenner-Alsop, 2009; Lotto & Holt, 2006). The uncertainty of the perception task in the V-set was also manifested in the larger width of the category boundary for both subject groups, in comparison to the boundary widths observed in the CVC-sets. In the categorization, the Estonian listeners obviously benefited from their language-specific tuning of duration processing, which resulted in the location of the category boundary at the mean duration of 139 ms (SD=6.2 ms). Somewhat surprisingly, a highly similar short/long boundary value (140 ms) was found in the perception experiment by Eek and Meister (2003) using CV(V)C-syllables whose temporal structure modeled a slow articulation rate. Moreover, this perceptual boundary value is only slightly longer than the maximum duration of a short vowel in the stressed syllable of a CVCV word in slow articulated speech (mean=129 ms, SD=18.1 ms) and close to the minimum duration of a long vowel in the stressed position of a CVVCV-word pronounced at a fast articulation rate (mean=136 ms, SD=21.8 ms) as measured in the production data (Eek & Meister, 2003). This leads to the hypothesis that for Estonian subjects there might be an absolute context-independent short/long category boundary, which in the case of vowels would seem to lie around 140 ms (more evidence is of course needed to test this hypothesis).

The category boundary of the L2 subjects lies at a slightly longer duration (mean=146 ms, SD=6.8 ms) than that of the

L1 subjects. The L2 subjects might have exploited a continuous auditory processing mechanism and, instead of a real category boundary, their results might rather mirror the response strategy induced by the experimental design (see Massaro, 1987; Schouten, Gerrits, & van Hessen, 2003). Yet, as suggested by Ylinen et al. (2006), the longer boundary value of the Russian subjects might be due to influence from their L1. The production experiment of Finnish pseudo-words embedded in sentences revealed that Russian subjects tend to produce stressed short vowels significantly longer than do native Finnish subjects. In the categorization task, L2 subjects may have used a typical Russian CVCV word as a reference, and they may have classified stimulus vowels as "long" when these vowels were much longer than their mental pattern of the L1 stressed vowel (Ylinen et al., 2006). In line with the production data on Finnish CVCV words, it has been observed (Meister & Meister, 2005) that in Russian speakers' pronunciation of Estonian CVCV words, the duration of the stressed-syllable vowel is up to 1.5 times longer than in native Estonian speakers' pronunciation. Thus, the interpretation provided by Ylinen and her colleagues for Finnish seems to apply to Estonian, too, and can be considered a likely explanation of the differences of category boundary values between the L1 and L2 groups in all of our stimulus sets.

However, since the difference in the category boundary between the L1 and L2 groups is rather small (albeit statistically significant), we cannot entirely dismiss the possibility that the L2 subjects' responses demonstrated vaguely established short/long categories acquired during L2 learning as well as in everyday L2 exposure and practice. On the other hand, it should be noted that the perception experiments using Finnish isolated vowels did not reveal any differences in the location of the category boundary nor in the consistency of responses between Russian L2 users and naïve Russians (Ylinen et al., 2006).

Due to the presence of phonetic context in the CVC-stimuli (durations of the abutting consonants), both subject groups had the short/long category boundary at much shorter vowel durations than in the V-set; that is, given the context, the categorization task was easier and more realistic despite the fact that the stimuli were pseudo-words in both languages. Both subject groups showed consistent responses, and they exhibited almost equal category boundary width in both stimulus sets.

The two CVC-subsets were designed to model moderate and fast speaking rates by using two different duration values for surrounding consonants ($C1,2=110$ ms for moderate and $C1,2=80$ ms for fast rate). The results of both subject groups demonstrated a similar pattern—the short/long vowel boundary occurs at shorter duration in the case of shorter consonant duration. The results are in line with an earlier study in which Estonian subjects showed a proportional relationship between the duration of a word-initial consonant and the perceived short/long category boundary of the following vowel (Eek & Meister, 2003). The difference in category boundary locations between the L1 and L2 subject groups in both CVC-subsets is almost equal to the difference found in the V-set, which suggests that the interpretations suggested above for the results of the V-set are also applicable to the CVC-subsets.

The results of the CVCV-sets provide different response patterns for L1 and L2 subjects. The CVCV stimulus sets were designed to elicit short/long judgments at the phonemic level in that only the duration of the primary-stressed vowel was manipulated under two conditions—CVCV(1) with natural F0-contour and CVCV(2) with monotonous F0. For L1 subjects the CVCV vs. CVVCV opposition, in natural speech, is not just a short/long opposition of the primary-stressed vowel (the only variable here manipulated), but it also unavoidably evokes the language-specific foot-level contrast of quantity degrees Q1 vs. Q2, in which the full set of accompanying features (duration ratio of first and second syllable, characteristic F0 curve, intensity differences of stressed and unstressed syllables) is at the disposal of native subjects (see above the description of Estonian quantity system). There are no similar contrasting patterns in Russian that L2 listeners could exploit in distinguishing Estonian Q1 and Q2. Consequently, L1 subjects can be expected to exhibit categorization patterns different from those of L2 subjects. And indeed, L1 subjects demonstrated higher consistency of responses, much narrower width of the category boundary, and a larger difference between the categorization curves of two stimulus sets in comparison to those of L2 subjects.

Fundamental frequency was observed to have an important effect on the location of the short/long category boundary for the native Estonian subjects. Thus the category boundary was located at a significantly shorter vowel duration in the CVCV(1) subset (with natural F0) than in CVCV(2) subset (with flat F0). This was an expected result as several studies (e.g. Eek & Meister, 1997; Lehiste, 1997, etc.) have shown the importance of F0 in the perception of Estonian quantity degrees.

The categorization strategy of the L2 subjects in the CVCV-sets can be expected to be different from that of L1 subjects as no comparable phonological patterns to rely on exists in Russian. And as our experimental results show, the most salient cue for L2 subjects concerning the CVCV vs. CVVCV opposition was the duration of the stressed syllable vowel whereas F0 had no significant effect on categorization. This is in line with the results by Lippus et al. (2009), reporting perception results of Estonian quantity degrees by native Estonian and L2 speakers with Finnish, Russian, and Latvian backgrounds. Lippus et al. found that, in the CV(V)CV word structure, the Russian (and also the Finnish) subjects perceived quantity oppositions on the basis of temporal structure only and that the pitch cue had no effect, thus demonstrating significant contrasts with Estonian subjects.

The question of whether the present results support the Feature Hypothesis by McAllister et al. (2002) or the "desensitization" hypothesis proposed by Bohn (1995), must be left open since our results do not directly favor either hypothesis. In fact, both hypotheses have been proposed to explain the role of L1 transfer in vowel perception in a cue weighting paradigm, i.e. when both spectral and duration cues are involved in the perception (and production) of non-native contrasts. This is not exactly the case in Estonian short/long categorization since, for the L2 subjects, this contrast was seen to be based on the duration cue only. Our results show that L2 subjects were quite successful in distinguishing the Estonian short/long duration categories despite the non-categorical use of the duration cue in their native language. However, it is not possible to say which perceptual strategy the L2 subjects exploited. It is impossible to decide whether they relied on the salient durational cue, utilizing the continuous auditory perception mechanism, or whether they relied on variable (non-contastive) durational patterns of their L1 in which duration acts as a major cue to stress. Or did the L2 listeners exploit a combination of both strategies? Nor can the learning effect due to frequent L2 exposure be underestimated.

Acknowledgments

This research was partially supported by the target-financed theme No. 0322709s06 of the Estonian Ministry of Education and Research and by the National Program for Estonian Language Technology. We are grateful to all subjects who participated in the study, to Associate Editor Ocke-Schwen Bohn, and three anonymous reviewers for their helpful comments. Our special thanks are due to Kari Suomi for his valuable suggestions and constructive remarks on an earlier draft of the paper.

References

Ariste, P. (1939). A quantitative language. In *Proceedings of the Third International Congress of Phonetic Sciences*, Ghent, pp. 276–280.

Asu, E. L., Lippus, P., Teras, P., & Tuisk, T. (2009). The realization of Estonian quantity characteristics in spontaneous speech. In M. Vainio, R. Aulanko, & O. Aaltonen (Eds.), *Nordic prosody. Proceedings of the Xth conference* (pp. 49–56). Frankfurt, Peter Lang Verlag.

Bohn, O.- S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 279–304). Baltimore: York Press.

Bolotova, O. (2003). On some acoustic features of spontaneous speech and reading in Russian (Quantitative and qualitative comparison methods). In *Proceedings of XV ICPhS*, Barcelona, pp. 913–916.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, from 〈http://www.praat.org/〉.

Bondarko, L. V. (1998). *Fonetika sovremennogo russkogo jazyka*. St.-Petersburg: University of St.-Petersburg.

Bondarko, L. V., Volskaya, N., Tananaiko, S., & Vasilieva, L. (2003). Phonetic properties of Russian spontaneous speech. In *Proceedings of XV ICPhS*, Barcelona, pp. 2973–2976.

Cebrian, J. (2006). Experience and the use of duration in the categorization of L2 vowels. *Journal of Phonetics*, *34*, 372–387.

De Silva, V. (1999). *Quantity and quality as universal and specific features of sound systems. Experimental phonetic research on interaction of Russian and Finnish sound systems*. Studia Philologica Jyväskyläensia 48. Jyväskylä: University of Jyväskylä.

De Silva, V., Iivonen, A., Bondarko, L. V., & Pols, L. (2003). Common and language dependent phonetic differences between read and spontaneous speech in Russian, Finnish and Dutch. In *Proceedings of XV ICPhS*, Barcelona, pp. 2977–2980.

Eek, A. (1974). Observations on the duration of some word structures: I. *Estonian papers in Phonetics*, 18–31.

Eek, A. (1980). Further information on the perception of Estonian quantity. *Estonian Papers in Phonetics*, *1979*, 31–57.

Eek, A. (1987). The perception of word stress: a comparison of Estonian and Russian. In R Channon, & L Shockey (Eds.), *In Honor of Ilse Lehiste – Ilse Lehiste Pühendusteos* (pp. 19–32). Dordrecht—Holland/Providence, USA: Foris Publications.

Eek, A., & Meister, E. (1997). Simple perception experiments on Estonian word prosody: Foot structure vs. segmental quantity. In I. Lehiste, & J. Ross (Eds.), *Estonian prosody: Papers from a symposium* (pp. 71–99). Tallinn: Institute of Estonian Language.

Eek, A., & Meister, E. (1998). Quality of standard Estonian vowels in stressed and unstressed syllables of the feet in three distinctive quantity degrees. *Linguistica Uralica*, *3*(*34*), 226–233.

Eek, A., & Meister, E. (2003). Foneetilisi katseid ja arutlusi kvntiteedi alalt (I). Häälikukestusi muutvad kontekstid ja välde. *Keel ja Kirjandus*, *11–12*. 815–837, 904–918.

Eek, A., & Meister, E. (2004). Foneetilisi katseid ja arutlusi kvantiteedi alalt (II) : Takt, silp ja välde. *Keel ja Kirjandus*, *47*(5), 336–357.

Hayes-Harb, R. (2005). Optimal L2 speech perception: Native speakers of English and Japanese consonant length contrasts. *Journal of Language and Linguistics*, *4*, 1–29.

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *The Journal of the Acoustical Society of America*, *116*, 2384–2394.

Kalvik, M.-L., & Mihkla, M. (2010). Modelling the temporal structure of Estonian speech. In I. Skadina, & A. Vasiljevs (Eds.), *Human Language Technologies: The Baltic Perspective: Proceedings of the fourth international conference, Baltic HLT 2010*, Vol. 219 (pp. 53–60). Amsterdam: IOS Press, Frontiers in Artificial Intelligence and Applications.

Karlsson, F. (1969). Suomen yleiskielen segmentaalifoneemien paradigma. *Virittäjä*, *73*, 351–362.

Kingston, J., Kawahara, S., Chambless, D., Mash, D., & Brenner-Alsop, E. (2009). Contextual effects on the perception of duration. *Journal of Phonetics*, *37*, 297–320.

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America*, *124*(6), 3959–3972.

Krull, D. (1991). Stability in some Estonian duration relations. In *Phonetic experimental research.* Institute of Linguistics, University of Stockholm (PERILUS), Vol. 13, pp. 57–60.

Krull, D. (1992). Temporal and tonal correlates to quantity in Estonian In *Phonetic experimental research*. Institute of Linguistics, University of Stockholm (PERILUS), Vol. 15, pp. 17–36.

Krull, D. (1993). Word-prosodic features in Estonian conversational speech: some preliminary results. In *Phonetic experimental research. Institute of Linguistics*. University of Stockholm (PERILUS), Vol. 17, pp. 45–54.

Krull, D. (1997). Prepausal lengthening in Estonian: Evidence from conversational speech. In I. Lehiste, & J. Ross (Eds.), *Estonian prosody: Papers from a symposium* (pp. 136–148). Tallinn: Institute of Estonian Language.

Krull, D., & Traunmüller, H. (2000). Perception of quantity in Estonian. *Proceedings, Fonetik 2000*, pp. 85–88.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *The Journal of the Acoustical Society of America*, *63*, 905–917.

Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. In *American Studies in Uralic Linguistics 1* (pp. 21–82), Bloomington, Indiana University Press.

Lehiste, I. (1997). Search for phonetic correlates in Estonian prosody. In I. Lehiste, & J. Ross (Eds.), *Estonian prosody: Papers from a symposium* (pp. 11–35). Tallinn: Institute of Estonian Language.

Lehiste, I. (2003). Prosodic change in progress: From quantity language to accent language. In P. Fikkert, & H. Jacobs (Eds.), *Development of prosodic systems* (pp. 47–65). Berlin, New York: Mouton de Gruyter.

Lehiste, I., & Fox, R. A. (1992). Perception of prominence by Estonian and English listeners. *Language and Speech*, *35*, 419–434.

Liiv, G. (1961). Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid. *Keel ja Kirjandus*, *412–424*, 480–490.

Lippus, P. (2010). Variation in vowel quality as a feature of Estonian quantity. In M. Hasegawa-Johnson, A. Bradlow, J. Cole, K. Livescu, J. Pierhumbert, & C. Shin (Eds.) *Speech prosody 2010*, Chicago, USA. 〈http://aune.lpl.univ-aix.fr/∼sprosig/sp2010/papers/100877.pdf〉.

Lippus, P., & Pajusalu, K. (2009). Regional variation in the perception of Estonian quantity. In M. Vainio, R. Aulanko, & O. Aaltonen (Eds.), *Nordic prosody—Proceedings of the Xth conference* (pp. 151–157). Frankfurt: Peter Lang Verlag.

Lippus, P., Pajusalu, K., & Allik, J. (2009). The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics*, *37*, 388–396.

Lotto, A., & Holt, L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*, *68*, 178–183.

Massaro, D. W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception* (pp. 254–283). Cambridge, MA: Harvard University Press.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*, 229–258.

Meister, L., & Meister, E. (2005). Acoustic correlates of Russian accent in Estonian. In G. Kokkinakis, N. Fakotakis, E. Dermatas, & R. Potapova (Eds.) *Proceedings of SPECOM 2005: 10th International Conference Speech and Computer* (pp. 437–440). Patras, University of Patras.

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Research*, *16*, 492–495.

Remijsen, B., & Gilley, L. (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics*, *36*(*2*), 318–344.

Skrelin, P. (2004), Segment features in different speech styles. In *Proceedings of SPECOM'2004: 9th International Conference Speech and Computer (pp. 11–16).* Saint-Petersburg, Russia.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(*1*), 71–80.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(*2*), 420–428.

Suomi, K., Toivanen, J., & Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics*, *31*(*1*), 113–138.

Suomi, K., & Ylitalo, R. (2004). On durational correlates of word stress in Finnish. *Journal of Phonetics*, *32*, 35–63.

Suomi, K. (2005). Temporal conspiracies for a tonal end: Segmental durations and accentual f0 movement in a quantity language. *Journal of Phonetics*, *33*, 291–309.

Suomi, K. (2006). Stress, accent and vowel durations in Finnish. In G. Ambrazaitis, & S. Schötz (Eds.), *Lund University, Centre for Languages and Literature, Department of Linguistics & Phonetics, Working papers*, Vol. *52*.

Suomi, K. (2007). On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics*, *35*, 40–55.

Suomi, K. (2009). Durational elasticity for accentual purposes in Northern Finnish. *Journal of Phonetics*, *37*, 397–416.

Suomi, K., Toivanen, J., & Ylitalo, R. (2008). *Finnish sound structure: Phonetics, phonology, phonotactics and prosody. Studia humaniora ouluensia 9*. Oulu University Press.

Traunmüller, H., & Krull, D. (2003). The effect of local speaking rate on the perception of quantity in Estonian. *Phonetica*, *60*, 187–207.

Volskaya, N., & Skrelin, P. (2004). Russian read and spontaneous speech: Prosodic data analysis. In *Nordic Prosody IX*, Lund.

Ylinen, S., Shestakova, A., Alku, P., & Huotilainen, M. (2005). The perception of phonological quantity based on durational cues by native speakers, second-language users and nonspeakers of Finnish. *Language and Speech*, *48*(*3*), 313–338.

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. *Brain Research*, *1072*, 175–185.