Original Paper

Phonetica 2003;60:187–207
DOI: 10.1159/000073502

Received: October 15, 2002
Accepted: June 23, 2003

The Effect of Local Speaking Rate on the Perception of Quantity in Estonian

Hartmut Traunmüller   Diana Krull

Department of Linguistics, Stockholm University, Stockholm, Sweden

Abstract

The Estonian language with its elaborate system of contrasts in quantity, whose essentials are described in the paper, is used to investigate human perception of distinctive contrasts in the duration of vowels, consonants and larger units. In the experiments reported, the speaking rate of a preceding or following syllable was manipulated in addition to that of a target V, C or VC sequence that carried a quantity distinction in disyllabic words. The results confirmed that the second syllable in such words, in particular the duration of its vowel, serves as a reference, but they showed segments of additional syllables to contribute in the same direction. The results provided no support for ascribing quantity to any larger units than phonetic segments. Speech rate effects of similar magnitude have been observed in Japanese, while effects of the same kind were found to be smaller in Dutch. These differences may be linked with the functions durational contrasts have in the different languages. It appears that listeners have to adapt more fully to variations in the local speaking rate when there are no additional cues and the functional load of quantity distinctions is high.

Introduction

The present investigation has the primary aim of contributing to our knowledge about how listeners evaluate linguistic information conveyed by segment durations. Although segment durations have some function in all languages, their functional load varies widely between languages. Since natives of languages in which segment durations have a high functional load are forced to pay close attention to segment durations, both as speakers and as listeners, we have chosen Estonian as our test language. It has been shown that duration contrasts are maintained more consistently by Finnish and Estonian speakers than by speakers of languages with a simpler quantity system, like that of Swedish [Engstrand and Krull, 1994], and it can be suspected that an analogous between-language difference exists in perception as well. Contributing to an understanding of the quantity system of Estonian was considered a secondary aim of the present investigation, but its exposition and discussion will require most of the space of this introduction.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2003 S. Karger AG, Basel
0031–8388/03/0603–0187
$19.50/0
Accessible online at:
www.karger.com/journals/pho

Hartmut Traunmüller
Institutionen för lingvistik
Stockholms universitet,
SE–106 91 Stockholm (Sweden)
Fax +46 8 15 53 89, E-Mail hartmut@ling.su.se.

Table 1. Oppositions in segmental quantity in stressed position

	V1	V2	V3
C3	VC::		V::C::
C2	VC:	V:C:	
C1	VC	V:C	V::C
C0	(V)	V:	V::

Rows: same consonant quantity; columns: same vowel quantity.

A central concept in the traditional description of quantity in Estonian is the notion of *sisehäälikud* or 'inner sounds'. These are the sounds from the beginning of the vowel of a stressed first syllable to the beginning of the vowel of the second syllable. The main stress of a word is always on the first syllable, except in certain interjections and recent loan words. Within the inner sounds, Estonian distinguishes three degrees of quantity, commonly referred to as 'short', 'long' and 'overlong' (or 'extra long'), in both vowels and consonants. In other positions, there are certain restrictions. This is not the place to describe all of these, but it is relevant in the present context that there is no quantity opposition in word-initial consonants or in the vowels of unstressed syllables. In the case of a nonshort consonant, the boundary between the first and the second syllable is generally considered to be located within it. A rhythmic foot consists of a stressed syllable that may be followed by one or two unstressed syllables. Any subsequent rhythmic feet in a word always carry some degree of stress on their first syllables. In monosyllabic feet, there is no opposition between long and overlong. The necessity of a second syllable in order for quantity to display all its possibilities has been interpreted as indicating that quantity is realized on a syllabic as well as a segmental level.

In Estonian, there is no opposition in voicing. In writing, the letters b, d, g are used to represent short stops while in intervocalic position p, t, k represent the long and pp, tt, kk the overlong ones. Nonshort vowels are represented by reduplicated letters, but the standard orthography lacks a means of indicating the distinction between long and overlong vowels. Table 1 shows the segmental quantity oppositions that occur in stressed position between a possible initial consonant and the rhyme of a following unstressed syllable. There is an alternative, suprasegmental way of viewing these oppositions. This is shown in table 2, where the VC sequences are arranged accordingly. In this table, the columns represent the quantity of the 'inner sounds'. Inner sounds with the same quantity can be differentiated according to the particular segments that carry the quantity distinction. This can be the vowel, the consonant or both, as represented by the rows in table 2.

Liiv [1962] measured the durations of vowels in stressed position and found these to be related as 1.00:1.72:2.02 (Q1:Q2:Q3) [cf. also Eek, 1974, 1975]. He also measured the durations of the unstressed vowels and confirmed that these were inversely related to those of the first (1.75:1.37:1.00). He suggested the ratio between the duration of the vowel in the stressed syllable and that of the vowel in the following unstressed syllable to be important for perception since these ratios are more different from each other, 0.73:1.60:2.58. He also described between-quantity differences in pitch pattern, but considered these as mere concomitant phonetic phenomena.

Table 2. A suprasegmental analysis of quantity oppositions in stressed position

	Q1	Q2	Q3
…C		VC:	VC::
VC	VC	V:C:	V::C::
V…		V:C	V::C
V	(V)	V:	V::

Rows: same carrier; columns: same inner sound quantity.

Lehiste [1960, 1965] expressed the opinion that Estonian quantity cannot be satisfactorily described as something that is associated with segmental phonemes exclusively. She motivated this by the existence of a nontemporal difference between Q2 and Q3, in pitch pattern, that is perceptually important when present, but also, and mainly, by the fact that the structure of a stressed syllable determines the phonetic length of a following unstressed syllable. Lehiste [1960, 1965] suggested the duration ratios between the first and the second *syllable* to be characteristic and essentially invariant for each quantity. However, since this claim was maintained alongside with the incompatible claim that the duration of syllable-initial consonants does not matter, it cannot have been assumed to be valid in sensu strictu. Actually, Lehiste [1960, 1975] took the duration of the vowel for the syllable duration. If the intervocalic consonant is short, the syllable boundary can be considered to be located at its beginning, and if it is long or overlong the boundary is located at an imaginary point inside it. The imaginariness of this point is a problem. Unless there is a cluster of different consonants instead of just one long or overlong, there is no objective way for a listener to know the syllable boundary. This is in itself not a valid argument against Lehiste's suggestion. Listeners might well imagine such boundaries, but it has never been shown that this occurs in perception.

In a perception experiment, Lehiste [1975, 1997] varied the duration of the first vowel or the following consonant in synthetic disyllabic stimuli and used three alternative durations of the second syllable vowel. She interpreted her results as confirming the syllabic ratio hypothesis. Before doing any experiments of our own, we reanalyzed her data and our analysis confirmed that the second syllable contributed substantially to quantity perception, but not as much as implied by the ratio hypothesis. The possible contribution of a syllable-initial consonant remained unknown. If the ratio hypothesis were strictly valid, the duration of the second vowel (or syllable) would have the same perceptual weight as that of the first. This is in contrast with the conclusion that its duration is less important than that of the first vowel, also expressed by Lehiste [e.g., Lehiste and Danforth, 1977].

In Tauli's [1966] phonological analysis, the Estonian duration phenomena are interpreted in terms of two oppositions: (1) *phonemic quantity* and (2) *syllabic weight of stress*. Both are binary oppositions. 'Phonemic quantity' (short or long) is the opposition that distinguishes V1 from V2 (and V3) and C1 from C2 (and C3). In fact, Tauli [1966] represented these oppositions phonologically with V(V) and C(C), rather than V(:) and C(:). This choice was motivated by the fact that clusters such as diphthongs as well as postvocalic consonant clusters are perceived as equivalent in quantity with phonemically long segments. 'Syllabic weight of stress' (light or heavy) is the opposition that distinguishes Q2 (and Q1) from Q3. Tauli [1966] claimed Q3 to be a syllabic phenomenon, and his argument was based on consideration of cases in which a VC sequence carries the quantity.

As compared with a ternary opposition in both V and C segments (table 1), Tauli's [1966] analysis has the advantage that it does not leave any unused slots in the system, and it also accounts for the fact that Q3 occurs only in stressed syllables. It eliminates the problem of deciding whether the 'ambiguously long' segments that occur in trisyllabic words should be considered as in Q2 or Q3 [Lehiste, 1997]. Moreover, it resolved the conflict with the Prague school [e.g. Trubetzkoy, 1939], according to which all phonological oppositions ought to be binary. Tauli's [1966] analysis can also be said to have some phonetic support since Q3 is, typically, realized with a characteristic intonation pattern that contributes to its perception. This intonation pattern is, however, not always present [Krull, 1993, 2000] and can be compensated for by some additional segment duration, as can be concluded reliably from Lehiste's [1975, 1989, 1997] experiments. Tauli [1966] criticized Liiv's [1962] and Lehiste's ratio theories for their reliance on concomitant phenomena that hold only in words with a certain structure. However, for words similar in structure, such ratios have been observed to be very stable also in spontaneous speech, in which variation in speaking rate is frequent [Engstrand and Krull, 1994].

Like Tauli [1966], Eek [1980, 1990] and Eek and Meister [1997] analyze Estonian quantities as binary oppositions of length: short (Q1) vs. long, and weight: light vs. heavy (Q3). Eek [1980, 1990] uses the term 'accent' instead of 'quantity' for the weight distinction, and maintains that purely durational distinctions can occur only in intervocalic consonants. In all other cases, additional acoustic characteristics – such as F_0 contour and energy distribution – are involved.

Eek and Meister's [1997] conclusions rely crucially on the results of an experiment by Fox and Lehiste [1989], in which native speakers of Estonian and English were asked to *compare the vowel duration ratios* in pairs of synthetic [papap] stimuli and to tell whether they were the same or different. The ratios used were 1:2, 2:3, 3:2 and 2:1. The experiment produced similar results for the two groups of subjects. It showed that the listeners were only able to perform the comparison in terms of whether the first vowel was longer than the second or vice versa, which appears to be insufficient for handling Estonian. The authors concluded that 'duration ratios alone are insufficient to convey the linguistic distinctions'. However, it was already known from previous experiments [Lehiste 1975, 1997] that listeners are capable of distinguishing all quantity distinctions on the basis of durational cues alone. According to our interpretation, the results of Fox and Lehiste [1989] suggest that the perception of vowel quantity is different from the task the subjects had in their experiment: comparing vowel duration ratios. In an informally reported experiment with 2 listeners, Eek and Meister [1997] observed that listeners could only distinguish between short and long but not between long and overlong quantity when the vowel of the second syllable was cut off.

Lehiste [1965, 1997] described the Estonian quantity system as having several levels: 'the first syllable of a polysyllabic word is in one of three contrastive quantities, which may be called syllabic quantity 1, 2 and 3, or short, long and overlong. Syllabic quantity is not the sum of the segmental quantities of its components. Several possible consonant and vowel quantity combinations yield the same syllabic quantity; …' (cf. the columns of table 2 and the distinctions within each column). Due to the occurrence restrictions of quantity distinctions, the rhythmic foot can be considered as their domain. Within a foot, the three degrees of quantity are characterized by certain syllable duration ratios. In a word-initial foot, three degrees of quantity are possible in both consonants and vowels, later in the word, quantity opposition is found only in consonants. Lehiste [1975, 1997] does not use different labels, but speaks of quantity (Q1, Q2, Q3) on both levels of the hierarchy.

If segment duration is to provide the main cue for quantity distinctions but also varies as a function of speaking rate and word length, then it is logically impossible to rely on intrinsic measurements of the duration of one segment alone. In order to distinguish the underlying factors, it is necessary to relate the duration of a target segment to the durations of neighboring segments, to the local speaking rate or, equivalently, to measure segment duration with a 'slave clock' whose pace adjusts itself to the speaking rate of the utterance listened to [Traunmüller, 1994; cf. also Summerfield and Haggard, 1972]. In narrow scope, the local speaking rate can be said to reflect the effects of word length and word or breath group boundaries as well as the global speaking rate. Accounts of Estonian quantity based on the duration ratio between the first and second syllable (or rhyme) imply that no other segments than those of the second syllable are informative about the local speaking rate. However, it may be that the apparent role of the second syllable just reflects a more general phenomenon of perceptual adaptation to the local speaking rate. A contribution of subsequent information that can be interpreted as 'backward perceptual normalization' has been observed also in the perception of vowel quantity in Dutch [Nooteboom et al., 1978]. If the phenomenon is of a general kind, it is likely that segments within a wider temporal window contribute, so the preceding segments should also influence perception in the same direction as the following, but probably to a smaller degree. Variation in the quantity of contextual segments is not expected to have such an effect, but only variation indicative of speaking rate.

The idea of a model equivalent to that involving the use of a slave clock has been put into question by Nooteboom [1981] on the basis of observations that showed perceptual quantity boundaries to be affected less than in proportion to speaking rate in Dutch and in investigations of similar phenomena in other languages. In such cases, the listeners can be said not to be fully successful in distinguishing the different factors that affect segment durations. However, an investigation of Japanese quantity perception by Fujisaki et al. [1975] showed undiminished effects. Even in Nooteboom's [1981] alternative suggestion, listeners evaluate the durations of neighboring segments in judging the quantity of a segment in the same way as in a model that considers the local speaking rate.

Van Dommelen [1999] reported a perception experiment in which Norwegian listeners had been asked to identify stimuli as *mate* ['ma:tə] or *matte* ['matə]. It was previously known that there is a complementary relationship between the duration of the stressed vowel and the following consonant, and that the duration of the latter contributes to quantity perception. Now, van Dommelen [1999] showed that the duration of the final schwa also contributes to the same effect as the consonant. At least qualitatively, this resembles the situation in Estonian.

In order to see in which direction and to what extent extrinsic segments contribute to the perception of quantity we ran a perception experiment (experiment 1) that was to test the relevance of duration and duration ratios of syllables, rhymes and vowels, of the foot and the local speaking rate. We used stimuli in which the speaking rate of a

Table 3. Segment durations in milliseconds in the original utterances with *saagi* (left) and *satu* (right)

[ja]	[s]	[a:]	[ki]	[ka]
180	142	254	308	
	155	251	275	
	153	214	230	410

[ja]	[sa]	[t:]	[u]	[ka]
155	210	216	181	
	242	203	167	
	237	169	102	413

short preceding or following context was manipulated in addition to the duration of a V or a C that carried a quantity distinction. In manipulating V, C was kept short, and in manipulating C, V was kept short. In a second experiment (experiment 2), the whole quantity space was scanned, including a set of stimuli in which V and C were manipulated together. While the second experiment allowed us to see the generally valid regularities, the first experiment allowed a higher precision by scanning in smaller steps.

Experiment 1

Method

Stimuli

The stimuli were obtained by manipulating the durations of selected sections of recordings of the words *saagi* [sa:ki] 'crop, catch' (gen.) and *satu* [sat:u] 'get into' (imp., 2nd p. sg.) produced by a female broadcaster. The sampling frequency used was 44.1 kHz. The words were produced in three different contexts: preceded by *ja* [ja] 'and', in isolation, and followed by *ka* [ka] 'also'. The utterances were spoken with list reading intonation. The words [sa:ki] and [sat:u] were chosen because [saki] 'jostle' (imp., 2nd p. sg.), [sa::ki] 'saw' (part.), as well as [satu] 'rain' and [sat::u] 'get into' (imp., 2nd p. sg., parallel form) are also common words. The durations of the relevant segments are shown in table 3.

The durations of the [a:] in [sa:ki] and of the [t:] in [sat:u] (occlusion + burst) were modified in steps of, nominally, a factor $2^{n/8}$ (= 8 equal steps for a doubling or halving of the duration, which corresponds to a 9.05% increase in duration for each step) with $-9 < n < 9$, and with 15 consecutive values of n used in each context. The duration of the [t:] was modified by shortening or lengthening the silent interval as much as required with an unmodified burst. The duration of the [a:] was modified by repeating or eliminating periods equally distributed across the duration of the vowel. In order to obtain phase-clean joints, a deviation within ± ½ pitch period from the nominal duration was tolerated. The durations of either those parts of the utterances that preceded or that followed the target vowels and consonants were modified similarly, allowing n to take the values –2, 0, and +2 (84.0, 100, 118.9% of the original duration, respectively).

Listeners and Procedure

The listening sessions took place in the Laboratory of Phonetics and Speech Technology at Tallinn Technical University. The listeners consisted of three groups of Estonian university students, 27 in all, who were paid for their participation. The stimuli, which had been recorded on a CD, were presented binaurally through headphones (Koss KTX Pro). There were 480 stimuli separated by 0.6 s of silence and arranged in 32 series with additional pauses of 1 s in between. Within each series, the duration of only one target segment, [a] or [t], was varied, while the series differed in the modification of the durations of the other parts of the utterance, in the presence of *ja-* and *-ka*, and in the direction of durational change in the target segment. The listeners were also given two breaks of about 1–2 min during the session. An initial series of 15 extra stimuli served as practice. Each stimulus was presented

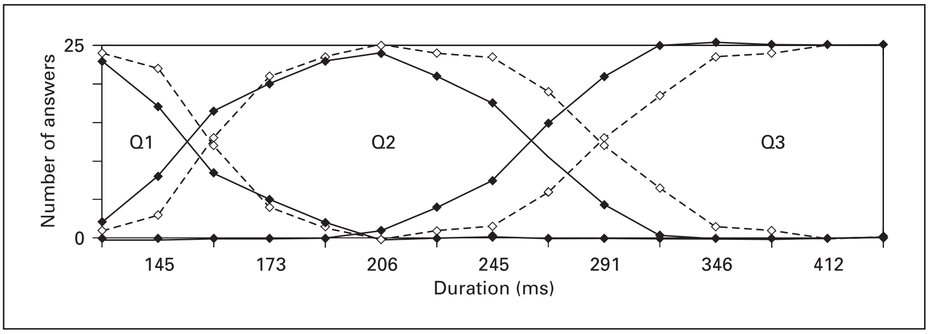

Fig. 1. Responses to stimuli derived from *ja saagi* with unmodified durations of ‘ja s’ and ‘gi’, plotted against the duration (logarithmically scaled) of the ‘aa’ for the two orders of stimulus presentation. Increasing duration: solid lines, decreasing duration: dashed lines.

once in a series with successively increasing duration of the target segment. Those beginning with *ja* were also presented in reverse order. The listeners’ task was to mark for each stimulus one or, in equivocal cases, two of the three words that were given on answer sheets for each series of stimuli, e.g. *sagi, saagi, s\`aagi* or *ja sadu, ja satu, ja sattu* (the accent sign in *s\`aagi* is one of the means of denoting quantity three where it is not clear from normal orthography).

In experiments of this kind, it is common to randomize the stimuli in order to blur the effects of presentation order in the pooled results. If this is to be efficient, a different randomization has to be used for each listener. We had several reasons not to choose this option, in addition to the fact that it would have been impractical: The method would have produced less sharp boundaries between quantity categories, which would result in less precise information on context effects, and it would not have shown the magnitude of the presentation order effect, which can be interesting in itself. In an attempt to reduce the context effects between successive series of stimuli, *saagi*-based series were always alternated with *satu*-based series and the series were arranged within groups of series characterized by the same modification of the context.

Results and Discussion

There were few response omissions. The responses from 2 subjects were excluded because they showed repeatedly inversions in the relation between segment length and quantity. Such inversions show that the subject’s decision criteria were unstable. Figure 1 shows the responses obtained from the remaining 25 listeners (13 male and 12 female) for vowel quantity in the stimuli derived from *ja saagi* when the durations of ‘ja s’ and ‘gi’ were unmodified. The figure shows for each of the 15 different durations of the vowel segment the number of answers obtained for each quantity. The data points that represent these numbers are connected by straight lines with their neighbors. We consider the boundaries between Q1, Q2 and Q3 to be located where the respective lines cross in the figure. In cases in which two response categories had been marked by a subject, the responses were shared between the categories.

In order to save space, we do not show all the results in the way we did in figure 1. Instead, we concentrate our interest on the quantity boundaries. The segment durations obtained for the quantity boundaries between Q1 and Q>1 and between Q3 and Q<3

Table 4. Quantity boundaries between Q1 and Q>1 and between Q3 and Q<3 of the vowel [a] observed with speech that had been manipulated in segment durations

Utterance	Rate structure	[ja]	[s]	[a] order 1	[a] order 2	[a:] order 1	[a:] order 2	[ki]	[ka]
ja saagi	> x >	151	119	*137*	*140*	*215*	*248*	259	
ja saagi	= x =	180	142	*153*	*158*	*261*	*290*	308	
ja saagi	< x <	214	169	*177*	*184*	*305*	*326*	366	
ja saagi	> x =	151	119	*149*	*154*	*239*	*264*	308	
ja saagi	= x =	180	142	*153*	*158*	*261*	*290*	308	
ja saagi	< x =	214	169	*162*	*167*	*256*	*331*	308	
saagi	> x =		130	*135*		*215*		275	
saagi	= x =		155	*134*		*227*		275	
saagi	< x =		184	*145*		*239*		275	
saagi ka	= x >		153	*124*		*204*		193	345
saagi ka	= x =		153	*141*		*226*		230	410
saagi ka	= x <		153	*148*		*257*		274	488

Durations of manipulated segments listed in milliseconds. Upright: given durations; in italics: durations at the quantity boundaries, in some cases for two orders of stimulus presentation: (1) successively increasing duration, (2) successively decreasing duration of the [a]. Rate structure: > increased, = unmodified, < decreased speaking rate (= increased segment durations) before or after x: the target vowel or consonant.

Table 5. Quantity boundaries for the consonant [t], otherwise analogous to table 4

Utterance	Rate structure	[ja]	[sa]	[t] order 1	[t] order 2	[t:] order 1	[t:] order 2	[u]	[ka]
ja satu	> x >	130	177	*124*	*128*	*204*	*242*	152	
ja satu	= x =	155	210	*146*	*150*	*237*	*282*	181	
ja satu	< x <	184	250	*157*	*166*	*276*	*306*	215	
ja satu	> x =	130	177	*113*	*131*	*216*	*260*	181	
ja satu	= x =	155	210	*146*	*150*	*237*	*282*	181	
ja satu	< x =	184	250	*148*	*155*	*256*	*288*	181	
satu	> x =		203	*119*		*241*		167	
satu	= x =		242	*131*		*248*		167	
satu	< x =		288	*150*		*265*		167	
satu ka	= x >		237	*120*		*211*		86	347
satu ka	= x =		237	*132*		*243*		102	413
satu ka	= x <		237	*140*		*271*		121	491

are shown in table 4 for the utterances in which vowel duration was varied and in table 5 for those in which consonant duration was varied. The boundaries were obtained by linear interpolation as illustrated in figure 1. There were never any Q1

Table 6. Percentage change of given and obtained segment durations (figures upright and in italics, respectively) at the boundaries listed in table 4

Utterance	[ja]	[s]	[a] order 1	[a] order 2	[a:] order 1	[a:] order 2	[ki]	[ka]
ja saagi	35	35	*26*	*28*	*34*	*27*	35	
ja saagi	35	35	*8*	*8*	*7*	*23*	0	
saagi		35	*7*		*11*		0	
saagi ka		0	*17*		*23*		35	35

Percentages defined as 100 (dur_{max} – dur_{min})/$dur_{unmodif}$.

Table 7. Percentage change of given and obtained segment durations at the boundaries listed in table 5 (analogous to table 3)

Utterance	[ja]	[sa]	[t] order 1	[t] order 2	[t:] order 1	[t:] order 2	[u]	[ka]
ja satu	35	35	*23*	*25*	*30*	*23*	35	
ja satu	35	35	*24*	*16*	*17*	*10*	0	
satu		35	*24*		*10*		0	
satu ka		0	*15*		*25*		35	35

responses at or beyond the boundary between Q2 and Q3, so that this boundary always agreed with that between Q3 and Q<3, and there was only one exception to the rule that there were no Q3 responses at the boundary between Q1 and Q2 (between Q1 and Q>1).

The boundary shifts induced by the manipulations of the segment durations (or speaking rate) of the parts of the utterance that preceded and followed the target segment are listed in tables 6 and 7 for the vowel [a] and the consonant [t], respectively. In these tables, the shift is expressed in terms of a percentage calculated as 100 (dur_{max} – dur_{min})/$dur_{unmodif}$. The shift is expressed in this way for the preceding and the following context as well as for the target segment. For the context, dur_{max} was equal to $2^{2/8}$ $dur_{unmodif}$, and dur_{min} equal to $2^{-2/8}$ $dur_{unmodif}$.

In table 6 as well as in table 7, the first row represents cases in which the change in speaking rate of the initial and final parts of the utterance was the same. In these cases one might expect boundary shifts in proportion to the change in the duration of the initial and final segments of the utterance (35%). The shifts actually obtained were smaller. Only the Q3 boundary in *ja saagi* in the series with successively increasing segment duration behaved in close agreement with this expectation. In table 6 it can be seen that the effect of modifying the duration of the following context [ki(ka)] without change in the preceding context is larger than that of modifying the duration of the preceding context [(ja)s] without change in the following. In table 7, it can be seen that the preceding [(ja)sa] had a substantial effect on the perceived quantity of the following [t],

Table 8. The effect of a small increase in the duration of the preceding and following context, expressed as a percentage of the effect of the same increase in the duration of the target segment

a Target segment: the vowel

Utterance		Q1			Q3	
	[(ja)s]	[a]	[gi(ka)]	[(ja)s]	[a]	[gi(ka)]
*ja s**aa**gi*	–25	+100	–38	–63	+100	–77
*s**aa**gi*	–24	+100		–61	+100	
*s**aa**gi ka*		+100	–10		+100	–24

b Target segment: the consonant

Utterance		Q1			Q3	
	[(ja)sa]	[t]	[u(ka)]	[(ja)sa]	[t]	[u(ka)]
*ja sa**t**u*	–25	+100	–56	–41	+100	–107
*sa**t**u*	–39	+100		–33	+100	
*sa**t**u ka*		+100	–12		+100	–31

notably at the Q1 boundary. This effect must be due mainly to the [a] that immediately precedes the [t], since the [(ja)s] shows a relatively small influence in table 6.

There is an alternative way of analyzing the effects of context and target segment. A small increase in the duration of the target segment has the effect of increasing the probability for a higher quantity to be perceived. This alternative approach involves calculating the effect of a small increase in the duration of the preceding and/or following context and expressing this effect as a percentage of the effect of the same increase in the duration of the target segment. The result of such an analysis is shown in table 8, which is based on presentation order 1.

Table 8 shows one pattern very clearly: The effects of durational changes in any other segments, preceding as well as following, are opposite to those of the target segments. There is, thus, no indication of any syllabic units involved in the perception of quantity. If the perception of quantity distinctions was based on the duration of syllables or the durational ratios of the syllables involved, then a lengthening of an initial consonant should have the same effect as an equal lengthening of the vowel. We would have a value of +100 for the [s] of *saagi* and we would expect positive values everywhere in the two [(ja)s] columns of table 8a. It has been observed before that an initial consonant does not actually have this effect. If the [(ja)s] was of no importance, we would have values of zero there. Instead, we see that an initial consonant does have some effect, albeit a small one, and it works in the direction opposite to that expected on the basis of a syllable-based hypothesis.

It is more important, however, that the duration of [(ja)sa], which includes the vowel [a] of the sequence of inner sounds in *satu*, affected the perceived quantity of [t] in the direction opposite to that expected if quantity was perceived on the basis of the duration of the [at] sequence. This is incompatible with a hypothesis based on dinstictiveness of rhyme duration (syllable minus initial consonant) as well as based on syllable duration.

The [gi(ka)] tails of the utterances with *saagi* work in the expected direction, although the effect was not as large as expected on the basis of a vowel duration ratio hypothesis. This can be seen most easily in table 6, where this hypothesis predicts all the values in the [a] and [a:] columns to agree with those in the [ki] column. Actually, we find values between 7 and 23% and between 17 and 34% where this hypothesis predicts 0 and 35%, respectively.

When we try to apply syllable-based hypotheses to intervocalic consonants as in *satu*, we encounter difficulties since the syllable boundary cannot be located with any greater precision than 'somewhere within the occlusion interval of the [t]'. However, the negative values in the [(ja)sa] columns of table 8b are distinctly incompatible with any hypotheses that assume quantity to be perceived on the basis of the duration of the first syllable or its rhyme. Each of these hypotheses predicts positive values everywhere in these columns.

The analysis represented by table 2 suggests the alternative of considering the [VC] sequence of inner sounds, whose duration is not so problematic to measure. However, the negative values in the [(ja)sa] columns are also incompatible with any hypotheses that assume quantity to be perceived on the basis of the duration of this sequence of sounds.

Our data suggest that in cases analogous to those at hand, quantity perception is based on the durations of vowels and consonants compared with the durations of segments in their vicinity that do not carry a quantity distinction and that listeners in this process attach a heavier weight to the following segments than to the preceding ones. In order to also investigate cases in which the quantity of a consonant is free to vary in addition to that of the vowel that precedes it, we did an additional experiment (experiment 2) in which it was attempted to cover the whole set of quantity combinations that occur in Estonian VC sequences.

As for the effect of presentation order, we obtained a mean shift of 5.0 ms of the Q1 boundary for [a] and of 8.4 ms for [t], while the effect on the Q3 boundary was substantially larger, 36.6 and 37.8 ms, respectively. This is substantially larger even if expressed as a percentage: 14.3 as against 3.2% for [a] and 15.9 as against 6.1% for [t]. In each series, the listeners reached the quantity boundaries earlier than in the series in which the same stimuli were presented in the opposite order. This can be understood as due to perceptual contrast or adaptation, but we can only offer a tentative explanation for the greater susceptibility of the Q3 boundary: It may be that longer segments in general are more susceptible to effects of this kind. Vowels in Q3 are typically, but not always, produced with a falling intonation, which is known to contribute to the perception of Q3 [Lehiste, 1997]. This was absent in our stimuli. However, the absence of this additional cue cannot be responsible here, since an effect of the same magnitude was obtained with [t:] as well as with [a:].

Experiment 2

Method

Stimuli

The stimuli were obtained by manipulating the durations of selected sections of recordings of the word *saate* [sa:t:e] 'get' (2nd p. pl.) preceded by *ja* [ja] 'and', in isolation, and followed by *ka* [ka] 'also', spoken with list reading intonation. This word had been chosen because the forms [sate], [sa:te], [sa::te], [sat:e], [sat::e] and [sa::t::e] also are more or less common existent

Table 9. Segment durations in milliseconds in the original utterances of experiment 2

[ja]	[s]	[a:]	[t:]	[e]	[ka]	[a:t:]	[sa:t:e]
160	105	176	171	198		347	650
	121	175	188	201		363	685
	96	137	152	101	378	289	486

Table 10. Segment durations in the original utterances of experiment 2, expressed in percent of the duration of [a:t:]

[ja]	[s]	[a:]	[t:]	[e]	[ka]	[sa:t:e]
46.1	30.3	50.7	49.3	57.1		187.3
	33.3	48.2	51.8	55.4		188.7
	33.2	47.4	52.6	34.9	130.8	168.2

words.[1] The speaker was the same as in experiment 1. The segment durations of the three original utterances are listed in tables 9 and 10.

A comparison of the relative segment durations shown in table 10 reveals a substantial final lengthening effect and also a minor effect of initial lengthening. These values are not necessarily representative of Estonian in general.

As in experiment 1, the durations of the [a:] and [t:] in [sa:t:e] (occlusion + burst) were modified in steps of, nominally, a factor $2^{n/8}$ with $-8 < n < 8$, but this time only the even or uneven values of n were used in each series of stimuli. This was done in order to reduce the number of stimuli to a number that the subjects could be expected to handle. As before, a deviation within ± ½ pitch period from the nominal duration was tolerated and neglected in the evaluation. The durations of either those parts of the utterances that preceded or that followed the [a:t:] sequence were modified similarly, allowing n to take the values –2, 0, and +2. Figure 2 illustrates the arrangement of the stimuli in the quantity space. The small circles represent the stimuli, which were arranged in seven series. In the three vertical series, the duration of [a] was kept constant at the durations shown on the x axis while [t] was lengthened step by step. In the horizontal series, the duration of [t] was kept constant while [a] was varied. In the diagonal series, [a] and [t] were covaried, keeping their relative durations constant. The constant durations were chosen so as to agree approximately with those observed in corresponding words produced by the same speaker in the same context at a normal speaking rate.

Listeners and Procedure

The listening sessions were carried out at the same place as in experiment 1. The listeners were 6 male and 21 female Estonian students (not the same as in experiment 1), divided into three groups. They were paid for their participation. The stimuli had been recorded on a CD and were presented through headphones (Koss KTX Pro). There were 1,020 stimuli separated by 0.6 s of silence and arranged in 154 series with additional pauses of 1 s in between. In addition, there was an initial series for practice. Within each series, the duration of the target segment [a], [t] or [at] was varied in one of

[1] The words [sate] and [satte] are not listed in the normative dictionary *ÕS 1999*. Therefore they were avoided in experiment 1. They are, however, listed in Saareste [1968].

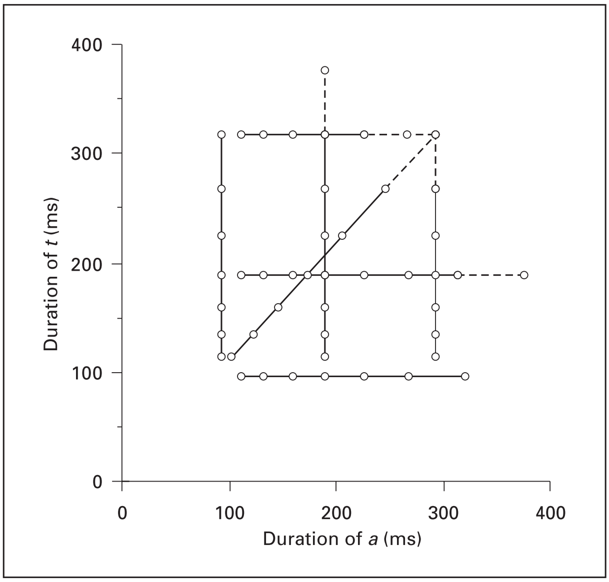

Fig. 2. The seven series of stimuli in quantity space. The duration values represent the word spoken in isolation. The longest durations (dashed lines) were only used when the final [e] or [e ka] was prolonged.

the seven ways shown in figure 2, while the series differed in the modification of the durations of the other parts of the utterance and in the presence of ja- and -ka. The listeners were also given three breaks of about 1–2 min during the session. Each stimulus was presented once in a series with successively increasing duration of the target segment ([V], [C], or [VC]). The listeners' task was to mark one or, in equivocal cases, two of the seven words that were given on the answer sheets for each series of stimuli, i.e. *sade, saade, s`aade, sate, satte, saate* and *s`aate*.

Results and Discussion

There were few response omissions, but the responses from 4 subjects (1 male and 3 female) were excluded because of unstable decision criteria that showed themselves in unpredictable inversions in the relation between segment length and quantity. In order to obtain an overview of the results obtained from the remaining 23 subjects, the boundaries between perceived Q1 and Q>1 and between Q<3 and Q3 were calculated in each series as in experiment 1, based on linear interpolation between the identifications obtained for the stimuli that were closest to the boundary. In a few cases, extrapolation was applied, but this was only allowed where the number of votes for the underrepresented alternative reached at least one third. In a few series, the number of Q1 responses was too low to obtain a Q1/ boundary.

Figure 3a shows the boundaries obtained in six of the seven ways (all but the diagonal series) for stimuli based on *saate* (without *ja* and *ka*) in which the duration of the initial [s] was modified (fast, normal and slow rate) while the final [e] retained its original duration. Figure 3b is analogous to 3a, but it shows the boundaries observed when the initial [s] retained its original duration while the duration of the final [e] was modified. It is evident that the duration of the final [e] has a more substantial effect on the location of the boundaries than that of the initial [s]. This holds for the quantity bound-

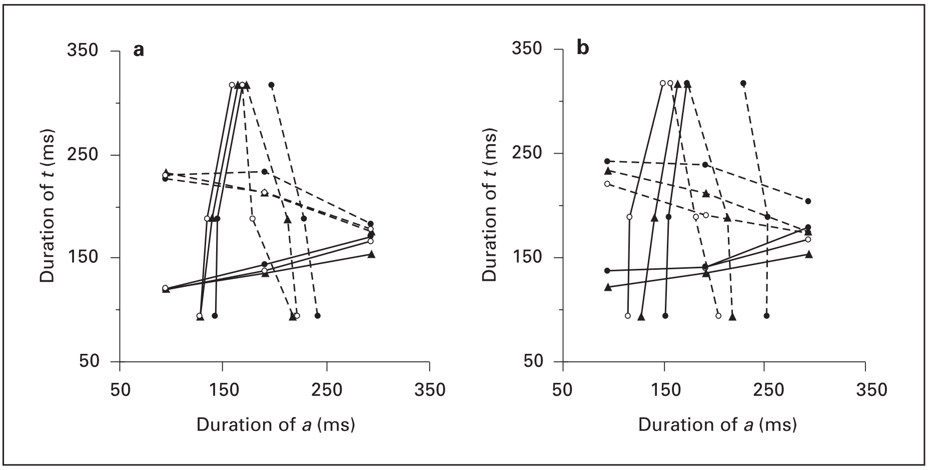

Fig. 3. Quantity boundaries observed in stimuli derived from [sa:t:e] (without [ja] or [ka]), presented with increased, unchanged and decreased rate of the initial [s] (**a**) and the final [e] (**b**). Solid lines: boundary between Q1 and Q>1. Broken lines: boundary between Q<3 and Q3. Unfilled circles: increased rate; triangles: unchanged rate; filled circles: decreased rate. To the left of the 'vertical' solid lines: [a] perceived as short by at least 50% of the listeners. To the right of the 'vertical' broken lines: [a] perceived as overlong. Below the 'horizontal' solid lines: [t] perceived as short. Above the 'horizontal' broken lines: [t] perceived as overlong.

aries of the consonant ([t]) as well as for those of the vowel ([a]). The /Q3 boundary of [a] appears to be most sensitive to the duration of segments in the immediate neighborhood of the 'inner sounds' while the effect on the Q1/ boundary of [t] is small and inconsistent. The difference in slope between the Q1/ and /Q3 boundaries for [a] as well as for [t], can be understood as due to the nonuse of V:C:: and V::C: in the system.

The boundary values obtained in the diagonal series in which both V and C were increasing in duration are listed in table 11 in the same way as the boundary values are listed in table 4 for experiment 1.

The boundary shifts induced by the manipulations of the segment durations (or speaking rate) of the parts of the utterance that preceded and followed the VC sequence are listed in table 12. The shift is expressed in terms of a percentage calculated as 100 (dur_{max} – dur_{min})/$dur_{unmodif}$. The shift is expressed in this way for the preceding and the following context as well as for the target segment. For the context, dur_{max} was equal to $2^{2/8}$ $dur_{unmodif}$, and dur_{min} equal to $2^{-2/8}$ $dur_{unmodif}$.

In the first row of each triplet, the change in speaking rate of the initial and final parts of the utterance was the same. As in experiment 1, the boundary shifts obtained in these cases were smaller than the durational change in the initial and final segments of the utterance (35%). It can be seen more clearly than in the results of experiment 1 that the effect of modifying the duration of a following context [e(ka)] is larger than that of modifying the duration of a preceding context [(ja)s]. The markedly low weight of the initial [s] in *saate ka* as compared with *(ja) saate* suggests that if there is a following context of two syllables, the preceding context loses most of its importance. This would, then, be compatible with a segment or syllable ratio hypothesis involving

Table 11. Quantity boundaries between Q1 and Q>1 and between Q3 and Q<3 of [a] and [t] obtained for the series of utterances in which the duration of [a] and [t] had been varied together

Utterance	Rate structure	[ja]	[s]	[a] Q1/	[a:] /Q3	[t] Q1/	[t:] /Q3	[e]	[ka]
saate	= x =		121	*128*	*193*	*114*	*207*	201	
saate	> x >		102	*114*	*172*	*?*	*184*	169	
saate	< x =		144	*152*	*225*	*131*	*242*	239	
saate	> x =		102	*130*	*195*	*110*	*209*	201	
saate	< x =		144	*135*	*210*	*120*	*226*	201	
saate	= x >		121	*115*	*175*	*?*	*188*	169	
saate	= x <		121	*146*	*220*	*131*	*236*	239	
jasaate	= x =	160	105	*124*	*193*	*119*	*188*	198	
jasaate	> x >	135	88	*115*	*171*	*104*	*166*	166	
jasaate	< x <	190	125	*143*	*222*	*130*	*216*	235	
jasaate	> x =	135	88	*124*	*193*	*117*	*187*	198	
jasaate	< x =	190	125	*130*	*209*	*122*	*203*	198	
jasaate	= x >	160	105	*117*	*176*	*106*	*171*	166	
jasaate	= x <	160	105	*154*	*225*	*133*	*218*	235	
saateka	= x =		104	*105*	*152*	*101*	*169*	101	378
saateka	> x >		87	*92*	*147*	*83*	*164*	85	318
saateka	< x <		124	*117*	*175*	*114*	*195*	120	450
saateka	> x =		87	*102*	*158*	*101*	*175*	101	378
saateka	< x =		124	*104*	*158*	*106*	*175*	101	378
saateka	= x >		104	*94*	*148*	*?*	*165*	85	318
saateka	= x <		104	*118*	*168*	*116*	*186*	120	450

Segment durations in milliseconds. Upright: given durations; in italics: obtained boundary values. Rate structure: > increased, = unmodified, < decreased speaking rate (= increased segment duration) before or after the [at] sequence.

Table 12. Percentage change of given and obtained segment durations (figures upright and in italics, respectively) at the boundaries listed in table 11

Utterance	[ja]	[s]	[a] Q1/	[a:] /Q3	[t] Q1/	[t:] /Q3	[e]	[ka]
saate		35	*30*	*27*	*?*	*28*	35	
saate		35	*4*	*8*	*9*	*8*	0	
saate		0	*24*	*23*	*?*	*23*	35	
jasaate	35	35	*23*	*30*	*22*	*27*	35	
jasaate	35	35	*5*	*8*	*4*	*9*	0	
jasaate	0	0	*30*	*25*	*23*	*25*	35	
saateka		35	*24*	*18*	*31*	*18*	35	35
saateka		35	*2*	*0*	*5*	*0*	0	0
saateka		0	*23*	*13*	*?*	*12*	35	35

Percentages defined as 100 $(dur_{max} - dur_{min})/dur_{unmodif}$. ?: Unknown because of missing boundary value.

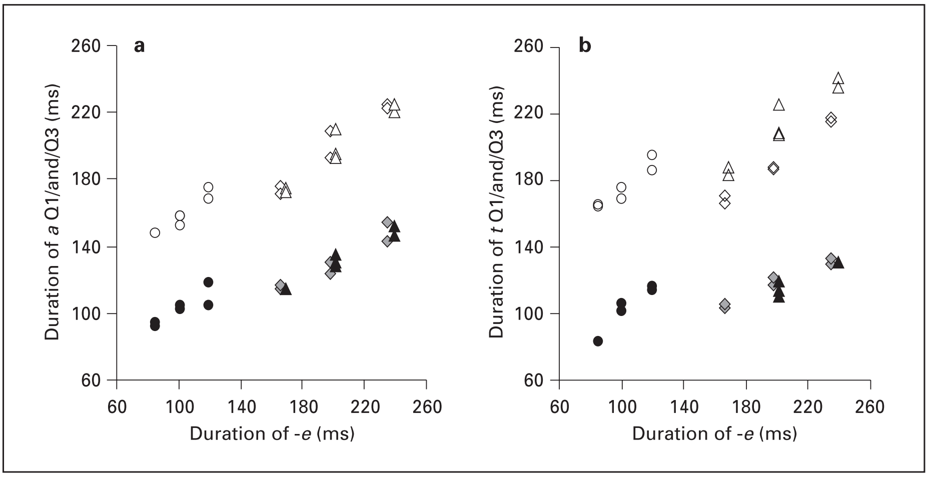

Fig. 4. Quantity boundaries Q1/Q>1 (filled symbols) and Q<3/Q3 (unfilled symbols) obtained for [a] (**a**) and for [t] (**b**) as a function of the duration of the [e] in the diagonal series, affected by final lengthening of the [e] in [sa:t:e] (triangles) and [ja sa:t:e] (squares) and unaffected in [sa:t:e ka] (circles).

the stressed and the post-stressed syllable, but the sensitivity to the number of syllables speaks against such a hypothesis.

Figure 4 shows the variation of the quantity boundaries obtained in the diagonal series as a function of the duration of the vowel [e] of the second syllable. If the syllable-ratio hypothesis was fully correct, the data displayed in figure 4 should fall on two linear regression lines – one for Q1/ and one for /Q3. In these data, we can, instead, see an effect of the presence or absence of a word or a syllable *(ka)* that follows the post-stressed one.

The data presented in table 13 reflect the result of a multidimensional regression analysis performed separately for each type of context (*ja s-e, s-e,* and *s-e ka*) with each of the two quantity boundaries as the dependent variables. In table 13 it can be seen that all significant effects of changes in the duration of any noninternal segments are opposite that of the same change in the duration of the internal target segment. However, the vowel of the second syllable has in each case a stronger effect than the initial consonant of the first syllable. At the /Q3 boundary of [a], a given increase (in milliseconds) has even a larger effect (108%) when it occurs on the vowel [e] of the second syllable. In calculating this percentage, it was assumed that the effect of a change in duration of the [e] alone would be equal to the effect actually observed with the change distributed uniformly across [e ka]. Since we cannot assure this, the calculated value must be seen as an upper limit. Even so, the value is lower than required in order for the perceived quantity of the [a] to be given by the duration ratio [a]/[e]. This is due to the shorter length of [e] in [e ka] as compared with the /Q3 boundary length of the vowel of the preceding syllable. There is an analogous discrepancy also at the Q1/ boundary and in the perceived quantity of the [t]. This confirms the results obtained in experiment 1 [the results obtained in experiment 1 were reported previously in Krull and Traunmüller, 2000].

Table 13. The effect of a small uniform increase in the duration of the [s] (or the [ja s]) that precedes the internal segments and of the [e] (or the [e ka]) that follows them, expressed as a percentage of the effect of an equally large increase in the duration of the internal target segment (bold), ± the standard error of the estimate

Utterance	Q1/					/Q3				
		(ja)s	a, t	e(ka)			(ja)s	a, t	e(ka)	
*ja s**aa**te*	?(**+0**±8)	?**+1**±20	**+100**	?**–34**±11		(**–15**±9)	**–37**±23	**+100**	**–64**±12	
*s**aa**te*		**–20**±19	**+100**	**–45**±12			**–39**±26	**+100**	**–69**±16	
*s**aa**te ka*		**–7**±18	**+100**	**–55**±19	(**–11**±4)		**–8**±30	**+100**	**–108**±31	(**–23**±7)
*ja saa**t**e*	(**–6**±14)	**–14**±35	**+100**	**–27**±18		(**–22**±11)	**–57**±28	**+100**	**–49**±15	
*saa**t**e*		?**–9**±29	**+100**	?**+1**±19			**–30**±24	**+100**	**–57**±15	
*saa**t**e ka*		**–12**±25	**+100**	**–31**±26	(**–7**±5)		**–2**±20	**+100**	**–96**±21	(**–20**±4)

The values are significant ($p < 5\%$) if this error is less than half the estimate. Results marked with ? are based on incomplete data and lack reliability.

It can be seen in this table that the effect of the initial [s] is mostly reduced when there is a final [ka]. Although statistical significance is not reached, this suggests that the pattern observed in table 12 and figure 4 is not specific to the diagonal series. This reduced weight of the initial [s] did not show itself in experiment 1 since the set of stimulus series was incomplete. In other essential respects, the data shown in table 13 are similar to those obtained in the previous experiment. The effect of a sole syllable-initial [s] failed to reach significance, and that of a preceding [ja s] reached significance only for consonant quantity at the /Q3 boundary. The effect of a following [e] or [e ka] reached significance in all cases at the /Q3 boundary, but at the Q1/ boundary only for the vowel.

The result of an overall linear regression analysis of all the boundary values obtained is shown in table 14. This analysis should be considered as tentative, since the duration of [ja] and [ka] had to be entered as zero when these syllables were absent, which probably does not correctly reflect how the listeners interpreted the stimuli. The figures entered in this table allow calculating the boundary values for the duration of the target segment. To give an example: for $dur_{[a]Q1/}$, the Q1/ boundary of [a], the equation is

$$dur_{[a]Q1/} = 26.6\text{ ms} - 0.02\,dur[ja] + 0.10\,dur[s] + 0.13\,dur[t] + 0.39\,dur[e] + 0.05\,dur[ka].$$

It can be seen in table 14 that listeners in each case attached some weight to the duration of the [e] of the post-stressed syllable, but substantially more at the /Q3 than at the Q1/ boundary (0.71 > 0.39 and 0.61 > 0.25). However, this difference in weight is not specific to the [e], but holds for the initial [s] and a finally attached [ka] as well. It can also be seen that the weight listeners attach to the [e] is higher in judgments of the duration of the vowel [a] of the stressed syllable than in judgments of the duration of the consonant [t] that follows it (0.39 > 0.20 and 0.71 > 0.61). This holds also for the attached [ka], but not for segments that precede the [at] sequence. The different polarities of the weights obtained at the Q1/ and /Q3 boundaries for [a] as well as for [t] reflect the different slopes of the boundary lines that can be seen in figure 3. The weight

Table 14. The result of an overall linear regression analysis: number of stimulus series n (of 84), correlation coefficient r and the unnormalized coefficients obtained for the durations of the listed segments

	a, Q1/	a, /Q3	t, Q1/	t, /Q3
n	77	84	78	83
r	0.945	0.951	0.971	0.932
C, ms	26.6	61.8	36.8	82.6
ja	(–0.02)	–0.04	(0.02)	(–0.02)
s	(0.10)	0.33	0.10	0.34
a			0.25	–0.19
t	0.13	–0.20		
e	0.39	0.71	0.20	0.61
ka	0.05	0.09	0.03	0.05

The duration of [ja] and [ka] had been set to zero when the syllable was not present. Values within parentheses: not significant.

of the introductory [ja] 'and' came out of this analysis as negligibly small, while that of the syllable-initial consonant [s] came out as half as high as that of the [e] for the consonant [t] and slightly less for the vowel [a].

General Discussion

The perception of quantity can be thought of as a process in which a listener evaluates the durations of speech segments of phonemic length or of any larger size by comparing their observed durations with a standard duration that is not constant but varies as an inverse function of the local speaking rate. The listener can be said to perform a division, and we can ask ourselves what is in the numerator of this division and what is in its denominator. The ratio hypothesis suggested by Lehiste [1960] can be interpreted in this way. According to this hypothesis, listeners judge the duration ratio between the first and the second syllable. Thus, listeners take the duration of the second syllable, which is in the denominator, as a unit of measurement that reflects the speaking rate. By and large, this sounds reasonable, but we have to be scrutinizing in confronting our results with the implications of this hypothesis.

Do we really have the duration of the first syllable in the numerator? Although Lehiste [1960] expressed her hypothesis in this way, she also claimed that the duration of the initial consonant of a syllable is not relevant. This would imply that she really meant 'rhyme duration' when she spoke of 'syllable duration'. We do not need to resolve this question, since the results of both of our experiments are incompatible with either interpretation. The minus signs and zeros that appear in tables 8 and 13 for the effects of any segments except the target segment show that we only have the duration of a single phone, either a vowel or a consonant, in the numerator. No other segment contributes to perceived quantity in the same direction, which would have been the case if quantity were perceived on the basis of any larger unit than a single segment. In this respect, the negative effect of the vowel [a] on the consonant [t], which was observed in experiment 1, is particularly important since it speaks against considering the rhyme of the first syllable as a perceptual unit.

Our second question concerns the denominator: Do we really have the duration of the second syllable (or its rhyme or just its vowel) in the denominator? The analysis presented in table 14, but also the results of experiment 1, show that the denominator contains more than the second syllable. It contains a weighted sum of segment durations that comprises approximately, but not strictly, a rhythmic foot. Within this range, the vowel of the second syllable has the largest weight. Basically, this is in agreement with the suggestion that the ratio between the duration of the vowel in the stressed syllable and that of the vowel in the following unstressed syllable is 'important' for perception [Liiv, 1962], which is also the essence of Lehiste's [1960] findings.

Although we have to reject the hypotheses that consider the durational ratio between the first and the second syllable of a rhythmic foot or between the vowels in these syllables as perceptual invariants, the present results do not give us any reason to question the presence of such invariance in speech [Engstrand and Krull, 1994]. Such invariant ratios are likely to result if speakers keep their speaking rate constant within a rhythmic foot. It appears reasonable to assume that this is usually the case, although the stimuli used in the present experiments did not stick to this pattern.

The susceptibility of quantity perception to the speaking rate that is evidenced in a following context is not unique to Estonian listeners. It has been observed in the perception of Dutch vowels and described as 'backward perceptual normalization' [Nooteboom et al., 1978]. English perception of [ba] vs. [wa], which involves segment durations more indirectly, has also been shown to be affected by later-occurring information [Miller and Liberman, 1979].

In the present investigation, the effect of context on the quantity distinction in the consonant was almost equally large as that on the vowel. The grossly similar magnitude of the effects is in line with Miller and Dexter's [1988] finding that English listeners use rate information from adjacent segments without regard to segment identity (vowel, approximant or stop) or phonotactics.

Our observations are similar to those made by van Dommelen [1999] on the effects of the duration of a vowel in the post-stressed syllable on the perception of quantity in the stressed syllable of the Norwegian word pair *mate* ['ma:tə] and *matte* ['matə]. Van Dommelen [1999] showed this to be a speech-specific process, which did not appear in non-speech stimuli consisting of tones and gaps with durations similar to those of the two vowels and the [t] of the speech stimuli used. The magnitude of the effects observed by van Dommelen [1999] was similar to that observed in our investigations at the Q1/Q2 boundary of the vowel, where the vowel duration was also similar to that at the boundary of the Norwegian pair. However, the variation in the duration of the unstressed vowel was substantially larger in the Norwegian study.

The effects of speech rate variation obtained in the present study are approximately as large as those observed by Fujisaki et al. [1975] with Japanese /isse/ vs. /ise/ presented in isolation. When these Japanese stimuli were presented in a carrier sentence, adaptation to speaking rate reached 100%. The effects observed with Dutch listeners [Nooteboom, 1981] and with English listeners who had to distinguish a sequence [pp] from simple [p] [Pickett and Decker, 1960] were smaller than 100%. However, the different studies are not immediately comparable and we cannot say anything conclusive about the nature of the between-language differences in perception. In particular, we cannot reject the possibility that the differences may be just quantitative in nature. They may be linked with the functions durational contrasts have in the different languages. In Estonian and Japanese, segment duration is distinctive for both vowels and consonants, and variables other than segment durations contribute only marginally to quantity distinctions. In Dutch, segment duration is distinctive only for vowels, in which it co-occurs with spectral cues. In English, segment duration has a more restricted role. In cases such as the one studied by Pickett and Decker [1960], speakers have the option of releasing the first stop to make it obvious that a doubble stop is meant. It appears that listeners have to adapt more fully to variations in the local speaking rate when there are no additional cues and the functional load of quantity distinctions is high.

If we allow for some bias towards a normal speaking rate, our results are compatible with the assumption that listeners relate the segment durations to a measure or an estimate of the local speaking rate or, equivalently and more directly, that they measure segment durations with a ‘slave clock’ whose pace adjusts itself to the local speaking rate. Moreover, the present data allow us to say that the local speaking rate is given by the durations of the segments within a window that holds a little more than two syllables. In quantity perception, this window appears to be centered approximately at the end of the segment whose quantity is judged. In table 12, it can be seen that the initial [s] lost nearly all of its importance when a final [ka] was appended. This suggests some mobility of the window. If this ‘slave clock’ principle is to work, it is necessary to assume some delay of the measurement, in order to allow the pace of the clock to be affected by a following syllable before the duration of a segment is judged. However, it remains for future investigations to assess the value of a slave clock in describing the behavior of listeners in whose native languages duration based distinctions are of less importance.

As for the question of the appropriateness of different ways of describing the quantity system of Estonian phonologically, we cannot draw any firm conclusions from our experiments. The results are compatible with the assumption that we are dealing with a ternary distinction in quantity, but they are also compatible with an analysis such as proposed by Tauli [1966] and Eek [1990], which assumes two different distinctive features to be involved. The fact that Q3 is typically realized with a characteristic F_0 pattern can be taken as an argument for a two-feature analysis, but against the background that this pattern is often unobservable, this is no strong argument. The observation that listeners need a second syllable in order to be able to distinguish Q3 from Q2 [Eek and Meister, 1997] can be understood on the basis of the phonological system of Estonian, which does not provide for this distinction to occur in monosyllables. A two-feature analysis allows us to ‘explain’ this property of the phonological system, but it remains as unclear as before in what the essential qualitative articulatory-acoustic difference between the two features would consist.

Acknowledgements

This research has been supported, in part, by grant F0558/1998 from HSFR, the Swedish Council for Research in the Humanities and Social Sciences. Our special thanks are due to Einar Meister for his help with the arrangement of the listening sessions at the Institute of Cybernetics, Tallinn Technical University. We are grateful to Patrice Speeter Beddor and an anonymous reviewer for valuable comments on an earlier draft of this paper.

References

Dommelen, W. A. van: Auditory accounts of temporal factors in the perception of Norwegian disyllables and speech analogs. J. Phon. *27:* 107–123 (1999).
Eek, A.: Observations on the durations of some word structures. I. Estonian Papers Phonet., pp. 18–37 (Academy of Sciences of the Estonian SSR, Tallinn 1974).
Eek, A.: Observations on the durations of some word structures. II. Estonian Papers Phonet., pp. 7–54 (Academy of Sciences of the Estonian SSR, Tallinn 1975).
Eek, A.: Further information on the perception of Estonian quantity. Estonian Papers Phonet., pp. 31–56 (Academy of Sciences of the Estonian SSR, Tallinn 1980).
Eek, A.: Units of temporal organization and word accents in Estonian. Linguistica Uralica *26:* 251–263 (1990).
Eek, A.; Meister, E.: Simple perception experiments on Estonian word prosody; in Lehiste, Ross, Estonian prosody: papers from a symposium, pp. 71–99 (Institute of Estonian Language and Authors, Tallinn 1997).
Engstrand, O.; Krull, D.: Durational correlates of quantity in Swedish, Finnish and Estonian: cross-language evidence for a theory of adaptive dispersion. Phonetica *51:* 80–91 (1994).
Fox, R.; Lehiste, I.: Discrimination of duration ratios in bisyllabic tokens by native English and Estonian listeners. J. Phon. *17:* 167–174 (1989).
Fujisaki, H.; Nakamura, K.; Imoto, T.: Auditory perception of duration of speech and non-speech stimuli; in Fant, Tatham, Auditory analysis and perception of speech, pp. 197–219 (Academic Press, London, 1975).
Krull, D.: Word-prosodic features in Estonian conversational speech: some preliminary results. Proc. ASCA Workshop on Prosody, Working Papers, vol. 41, pp. 148–151 (Department of Linguistics, Lund University 1993).
Krull, D.: Perception of Estonian word prosody in whispered speech, in van Dommelen, Fretheim, Nordic prosody. Proc. 8th Conf., pp. 153–164 (Lang, Frankfurt am Main, 2000).
Krull, D.; Traunmüller, H.: Perception of quantity in Estonian. Proc. Fonetik 2000, pp. 85–88 (Department of Languages, University of Skövde, 2000).
Lehiste, I.: Segmental and syllabic quantity in Estonian. Am. Stud. Uralic Ling., vol. 1 (Indiana University, Bloomington 1960).
Lehiste, I.: The function of length in Finnish and Estonian. Language *41:* 447–456 (1965).
Lehiste, I.: Experiments with synthetic speech concerning quantity in Estonian; in Hallap, Congressus Tertius Internationalis Fenno-Ugristarum Tallinnae habitus 17–23, VIII, 1970, pars I: Acta linguistica, pp. 254–269 (Valgus, Tallinn 1975).
Lehiste, I.: Current debates concerning Estonian quantity. FUSAC ’88, Proc. 6th Ann. Meet. Fenno-Ugric Stud. Ass. Canada, pp. 77–86 (University Press of America, Lanham 1989).
Lehiste, I.: Search for phonetic correlates in Estonian prosody; in Lehiste, Ross, Estonian prosody: papers from a symposium, pp. 11–33 (Institute of Estonian Language, Tallinn 1997).
Lehiste, I.; Danforth, D.G.: Foneettisten vihjeiden hierarkia viron kvantiteetin havaitsemisessa (Summary: The hierarchy of phonetic cues in the perception of Estonian quantity). Eripainos Virittajasta *4:* 404–411 (1977).
Liiv, G.: On the quantity and quality of Estonian vowels of three phonological degrees of length. Proc. 4th Int. Congr. Phonet. Sci., pp. 683–687 (The Hague 1962).
Miller, J.L.; Dexter, R.R.: Effects of speaking rate and lexical status on phonetic perception. J. exp. Psychol. hum. Percept. Perform. *14:* 369–378 (1988).
Miller, J.L.; Liberman, A.M.: Some effects of later-occurring information on the perception of stop consonant and semivowel. Percept. Psychophys. *25:* 457–465 (1979).
Nooteboom, S.G.: Speech rate and segmental perception or the role of words in phoneme identification; in Myers, Laver, Anderson, The cognitive representation of speech, pp. 143–150 (North-Holland, Amsterdam 1981).
Nooteboom, S.G.; Brokx, J.P.L.; de Rooij, J.J.: Contributions of prosody to speech perception; in Levelt, Flores d’Arcais: Studies in the perception of language, pp. 75–107 (Wiley, Chichester 1978).
ÕS 1999: eesti keele sõnaraamat. Erelt, T. ed. (Eesti Keele Sihtasutus, Tallinn 1999).
Pickett, J.M.; Decker, L.R.: Time factors in the perception of a double consonant. Lang. Speech *3:* 11–17 (1960).
Saareste, A.: Eesti keele mõisteline sõnaraamat (Dictionnaire analogique de la langue estonienne), vol. 2 (Vaba Eesti, Stockholm 1968), accessible on the Web: http://www.eki.ee/dict/saareste/.
Summerfield, Q.; Haggard, M.P.: Speech rate effects in perception of voicing. Speech Synthesis Percept., Rep. Speech Res. Progr., vol. 6, pp. 1–12 (Department of Psychology, The Queen’s University of Belfast 1972).
Tauli, V.: On quantity and stress in Estonian. Acta ling. Hafnensia *9:* 145–162 (1966).
Traunmüller, H.: Conventional, biological, and environmental factors in speech communication: a modulation theory. Phonetica *51:* 170–183 (1994).
Trubetzkoy, N.S.: Grundzüge der Phonologie. Travaux du cercle linguistique de Prague 7 (1939).