LINGUISTICA URALICA XLV 2009 3 doi:10.3176/lu.2009.3.01

EINAR MEISTER (Tallinn), *STEFAN WERNER* (Joensuu)

DURATION AFFECTS VOWEL PERCEPTION IN ESTONIAN AND FINNISH

Abstract. Identification of vowels in quantity languages is usually considered to be independent of vowel duration since duration is used to realise the quantity oppositions and thus supposed to not be available as a cue for other features. To test the role of microdurational variations in vowel category perception in Estonian and Finnish listening experiments with synthetic stimuli were carried out, involving five vowel pairs along the close-open axis. The results show that in the case of high-mid vowel pairs vowel openness correlates positively with stimulus duration; in mid-low vowel pairs such correlation was only found for some of the Finnish subjects. We explain the observed difference between high-mid and mid-low pairs with the hypothesis that in case of shorter perceptual distances in vowel quality (high-mid area of vowel space) intrinsic duration plays the role of a secondary feature to enhance perceptual contrast between vowels, whereas in case of mid-low oppositions the perceptual distance is large enough to guarantee the necessary perceptual contrast by spectral features alone and vowel intrinsic duration as an additional cue is not needed.

Keywords: Estonian, Finnish, vowel perception, intrinsic duration, category boundary.

1. Introduction

The prosodic structure of speech in general is determined by two components, one of which is subject to conscious speaker control, while the other is determined by physiological constraints of the human articulatory system. The first component realises so-called extrinsic features manifested by phonetic context and word and utterance-level prosody, the latter intrinsic features conditioned by inherent characteristics and constraints of the speech production mechanism. Extrinsic features are language specific and thus acquired in learning, whereas intrinsic features are considered to be independent from linguistic context and regarded as phonetic universals.

Experimental studies of microprosodic/intrinsic features of several non-quantity languages (see e.g. Black 1949; Peterson, Lehiste 1960; Di Cristo 1978) have shown that open vowels tend to have lower F0, higher intensity and longer duration than close vowels. Similar phenomena have been found also in the case of quantity languages, at least for Estonian and Finnish (Meister, Werner 2006, among others).

But in quantity languages like Estonian and Finnish vowel quality perception is expected to be unrelated to vowel duration since this parameter is exploited in the realisation of phonemic duration oppositions. In Finnish most segments can occur in a short and a long quantity, e.g. (only one meaning is given for homonyms): *tuli* /tuli/ 'fire', nom.sg.; *tuuli* /tu:li/ 'wind', nom.sg.; *tulli* /tul:i/ 'customs', nom.sg. or *sata* /sata/ 'hundred', nom.sg.; *sataa* /sata:/ 'hundred', part.sg.; *saattaa* /sa:t:a:/ 'accompany', 3.sg.; *saata* /sa:ta/ 'accompany', 2.sg.imperat. In Estonian the quantity system is even more complex: on the phonemic level a short vs. long opposition exists, on the foot level a three-way quantity contrast is possible, traditionally referred to as short (Q1), long (Q2) and overlong (Q3) quantity degrees, e.g.: Q1 *kalu* /kalu/ 'fish', part.pl.; Q2 *kaalu* /kaalu/ 'weight', gen.sg.; Q3 *kaalu* /kaa:lu/ 'weight', part.sg.; Q1 *kala* /kala/ 'fish', nom.sg.; Q2 *kalla* /kalla/ 'arum', nom.sg.; Q3 *kalla* /kal:la/ 'pour', 2.sg. imperat. (Eek, Meister 1999). Estonian quantity degrees are two-syllable prosodic units with distinct durational patterns based on various combinations of duration ratios of foot-internal neighbouring phonemes (Eek, Meister 2003). As the complementary cues, the location of the fundamental frequency peak and the distribution of acoustic energy in the foot contribute to the perceptual distinction of quantity degrees (Lehiste 1989; Eek 1994). The acoustic measurements of Estonian vowel formants in CV(V)CV context show that quality differences of the stressed vowels in Q1, Q2 and Q3 are small and do not exceed 1 Bark in F1—F2 vowel space (Eek, Meister 1998). Also in Finnish, formant frequency differences between long and short versions of the same vowel always lie below 1 Bark, even when comparing stressed and unstressed vowels (Wiik 1965).

The role of intrinsic features in speech perception has not been studied very extensively since according to the traditional approach higher order suprasegmental features "override" them in natural speech. One of the few studies addressing the perceptual role of intrinsic duration of vowels (Kouznetsov 2001) has shown that vowel duration is consciously controlled by Russian speakers to increase phonetic contrast in cases where other parameters, for example spectral parameters, lose their distinctive power. Similar findings have been reported for Spanish and Catalan (Solé 2007; Solé, Ohala [to appear]) where intrinsic vowel duration is found to be under control of a speaker to enhance phonetic contrast between different phonological categories.

In quantity languages duration has a different function than in non-quantity languages — it acts as the main feature of quantity oppositions and not as e.g. a major prominence-inducing factor. This function of duration leads one to anticipate no relationship between vowel duration and quality perception or at most a much weaker relationship than in non-quantity languages. Nevertheless, the fact that intrinsic duration of vowels is registered in both quantity and non-quantity languages, has lead us to the hypothesis that the contribution of intrinsic duration to vowel category perception could be universal.

In our pilot study (Werner, Meister 2008) we investigated both Estonian and Finnish exemplars of the /i/-/e/ vowel pair; the results showed a clear effect of segment duration on vowel perception: the longer the duration of the stimulus, the more often it was identified as /e/. In a further

experiment involving five vowel pairs along the close-open axis /i/-/e/, /y/-/ø/, /u/-/o/, /e/-/æ/ and /o/-/ɑ/, four Estonian listeners showed considerable between-subject variation; nevertheless, the overall tendency — perceived vowel openness correlates positively with stimulus duration — was obvious (Meister, Meister, Werner 2008).

In our present study we attempt to confirm the outcome of our earlier experiments with a larger number of subjects and to provide more reliable evidence for the role of microdurational variations in vowel category perception in both Estonian and Finnish. The paper is organised as follows: first, we will give an overview of the previous work on intrinsic duration, then the two-stage experimental setup will be introduced; next, the results of the perception tests will be presented, and finally, discussion and summary are provided.

2. Earlier work on intrinsic duration

Intrinsic (or micro)prosody in general and intrinsic duration in particular have been studied experimentally to some extent for several languages. Over decades, a number of papers have established rank orders of intrinsic vowel duration from longer low to shorter high vowels in different languages, some examples are summarised here (please note that duration-based ranking was not in all cases the main point of these papers):

- Peterson, Lehiste 1960 for English: æ → ɔ → u → ɑ → i → ə → ɛ → ʊ → ɪ
- Neweklowsky 1975 for German (similar also in Antoniadis, Strube 1984): a → ɔ → ɛ → œ → ʏ → ɪ → ʊ (speaker 1), and a → œ → ɛ → ɔ → ɪ → ʊ → ʏ (speaker 2)
- Van den Heuvel, Rietveld, Cranen 1994 for Dutch: a → i/u
- Eek, Meister 1998 for Estonian:
  - quantity degree 1: ø → æ → ɑ → y → ɤ → o/u → e → i
  - quantity degree 2: æ → y/o → ɑ → u → ɤ/e → ø/i
  - quantity degree 3: æ → ɑ → o → e → y → i → ø → u → ɤ
- Meister, Werner 2006 for Estonian and Finnish: low → mid → high

Although there are variations of the precise sequential order between and also within languages the general trend remains recognisable in most of the reported data.

As to the actual amount of durational differences, A. Di Cristo, in his comprehensive two-volume analysis of French micro- and macroprosody (Di Cristo 1978), reports intrinsic vowel duration ratios, computed from the mean duration of /a/, divided by the mean pooled durations of /i/ and /u/ and expressed as the percentage exceeding 100%, for French and other languages. The comparison ratios for low versus high vowels range from 16.5% (Estonian) to around 40% (Japanese, British English). His own results for French, with a typical ratio of approximately 20%, differ from earlier much higher ones proposed by M. Rossi (1972), ca. 40%. It has to be noted here, though, that A. Di Cristo analysed vowels in systematically varied contexts which does not hold true for all other intrinsic duration studies. A. Di Cristo also found evidence for much smaller inter-subject

duration variation than variation in intrinsic f0. For longer stretches of spontaneous speech, on the other hand, intrinsic duration variation between speakers often is considerably large, due to the effect of varying individual articulation rates, as a more recent study shows (Van den Heuvel, Rietveld, Cranen 1994).

What is shared by all previous work on intrinsic duration is the view of intrinsic vowel duration as being negatively correlated with vowel height. Despite minor differences in the observed rank orders the overall trend seems to be clear and empirically attested for a variety of languages. The most common explanation offered for the manifest trend is that the effort for raising of the tongue shortens the articulation of higher vowels (e.g. Ladefoged 1964).

The role of intrinsic features in speech perception is addressed only in few studies involving non-quantity languages. A study on Russian (Kouznetsov 2001) has shown that vowel duration contributes to categorisation of ambiguous vowel quality in cases where spectral parameters lose their distinctive power. The acoustic measurements (Kouznetsov, Ott 1987) show that the frequencies of the first two formants of Russian [i] and [e] are located in partly overlapping areas; the results of perceptual experiments also demonstrate auditory closeness of Russian [i] and [e]. The results of the experiments with synthetic vowels in CVC context demonstrate that in the case of ambiguous vowel quality with formant frequencies in the overlapping area of vowels [i] and [e], the frequency of stimulus perception as [i] is inversely proportional to vowel duration (Kouznetsov 2001).

Intrinsic duration plays a role only in the close-open (F1) dimension. A study on the perception of Finnish [i]-[y] vowels demonstrates that duration has no role in distinguishing vowel categories along the front-back (F2) dimension (Eerola, Savela, Laaksonen, Aaltonen 2002). Thus, we will concentrate on the relations between vowel duration and category perception in the F1 domain alone.

Contrary to the conception of intrinsic features as mechanical, there are claims that intrinsic vowel duration is linguistically specified and actively controlled by the speaker: "The results suggest that in Catalan and American English differences in vowel duration which correlate with vowel height distinctions are actively manipulated by the speaker and exploited as a cue to vowel height differences." (Solé, Ohala [to appear]).

3. Experimental setup

Estonian and Finnish vowel systems are matching well except for the non-low back vowel /ɤ/ occurring only in Estonian. As the acoustic quality of Estonian and Finnish counterparts is very close, it is justified to use one stimuli set for both languages. Five vowel pairs representing quality oppositions in the close-open dimension in both Estonian and Finnish were chosen for investigation: three high-mid vowel pairs (/i/-/e/, /y/-/ø/, /u/-/o/) and two mid-low vowel pairs (/e/-/æ/, /o/-/ɑ/).

Our two-stage experimental setup involves a pre-test, where the perceptual category boundary of the two vowels in each pair has to be found, and the main test, where three to four ambiguous formant

structures at category boundaries with varied duration were presented to listeners.

Ten native Estonian (5 male and 5 female) and ten native Finnish (4 male and 6 female) adults participated in the perception experiments voluntarily; none of the subjects were phonetically trained or reported any hearing problems.

3.1 Pre-test

The pre-test was designed to discover the category boundary area for each close-open vowel pair, as the intrinsic vowel duration is expected to play a role specifically in vowel categorization with stimuli of ambiguous quality. Ten Estonian subjects participated in this pre-test. Relying on the extensive similarity between the systems of non-central vowels in Estonian and Finnish we only informally verified the compatibility of averaged category boundaries at this stage with one additional Finnish pre-test subject.

3.1.1 Stimuli and procedure

The stimulus corpus was created by interpolating stepwise between prototypical values of the first three formants for each vowel pair and calculating evenly spaced steps through the F1/F2/F3 continua. The respective seventeen to nineteen gradually changing stimuli were synthesized with KlattWorks (McMurray in preparation), an implementation of a Klatt-type formant synthesizer (Klatt 1980). The formant values of prototype vowels (see Table 2) are approximations of stressed-syllable vowel formants taken from an evaluation of the Estonian BABEL database (Eek, Meister 1999). All stimuli used in the first experiment were synthesized with constant F0 of 100 Hz and a duration of 160 ms.

All testing was conducted in a sound-isolated room and stimuli were presented to subjects via high-quality headphones. The test was administered with Praat's (Boersma, Weenink 2008) multiple forced-choice test facility; each vowel was repeated three times in random order with no replay option. In the test listeners had to decide on vowel quality in a binary identification task with non-primed single stimuli. In total 264 stimuli (17—19 stimuli x 3 repetitions x 5 vowel pairs) were presented to subjects. The duration of the test turned out to be 20—25 minutes.

3.1.2 Results

The pre-test showed that the width and location of the category boundary area vary among subjects and vowel pairs, whereas the location of individual boundaries varies even more spanning 4 to 5 stimuli steps in different continua. The category boundaries of the Finnish subject showed similar variability as in the case of Estonian which further encouraged us to use the Estonian boundary data for both subject groups in the main experiment (and save time for additional tests).

On the basis of the pre-test results two stimulus sets were derived (see description below).

Table 1

Stimuli numbers derived from individual vowel category boundaries of Estonian subjects (S1...S10) and stimuli numbers of Finnish set

Vowel pair	Subject-specific stimuli for Estonian group										Stimulus set for Finnish group
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	
/i/-/e/	10–12	10–12	9–11	9–11	9–11	10–12	9–11	10–12	10–12	9–11	9–12
/y/-/ø/	9–11	8–10	7–9	8–10	8–10	9–11	8–10	8–10	8–10	9–11	8–11
/u/-/o/	9–11	9–11	8–10	10–12	10–12	9–11	9–11	10–12	10–12	9–11	9–12
/e/-/æ/	10–12	10–12	10–12	9–11	9–11	9–11	10–12	10–12	10–12	10–12	9–12
/o/-/ɑ/	8–10	9–11	10–12	9–11	9–11	9–11	9–11	9–11	9–11	9–11	8–11

Table 2

Formant frequencies (in Hz) of vowel prototypes and ambiguous stimuli

	Prototypes	Ambiguous stimuli					Prototypes
/i/-/e/	/i/	9	10	11	12		/e/
F1	250	330	340	350	360		400
F2	2205	2076	2058	2040	2022		1950
F3	3000	2776	2748	2720	2720		2580
/y/-/ø/	/y/	7	8	9	10	11	/ø/
F1	260	316	326	335	344	354	410
F2	1750	1675	1663	1650	1638	1625	1550
F3	2160	2194	2199	2205	2211	2216	2250
/u/-/o/	/u/	8	9	10	11	12	/o/
F1	300	366	375	384	394	394	450
F2	660	721	730	739	748	748	800
F3	2220	2288	2300	2313	2325	2325	2400
/e/-/æ/	/e/	9	10	11	12		/æ/
F1	400	527	543	559	575		670
F2	1950	1762	1738	1715	1691		1550
F3	2580	2495	2485	2474	2464		2400
/o/-/ɑ/	/o/	8	9	10	11	12	/ɑ/
F1	450	538	550	563	575	588	650
F2	800	888	900	913	925	938	1000
F3	2400	2426	2430	2434	2438	2441	2460

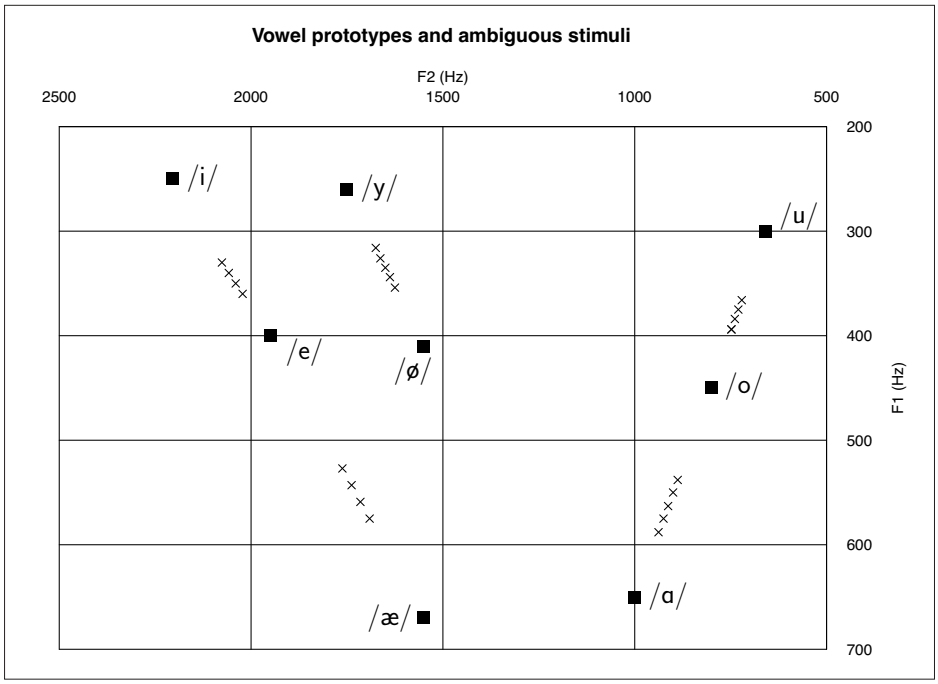

Figure 1. **Vowel prototypes (■) and ambiguous stimuli (+) in used in the main test.**

3.2 Main test

The main test was designed to answer the primary question of the study — do microdurational variations affect the perception of vowel quality in quantity languages?

3.2.1 Stimulus sets

Two different approaches were implemented in the preparation of the stimulus sets: (1) stimulus set based on subject-specific category boundaries, (2) stimulus set based on the averaged category boundaries.
(1) The pre-test results showed individual variations of the category boundary location, while the width of the boundary areas was in most cases limited to 1—2 contiguous stimuli. For the subject-specific stimulus set three stimuli around the individual category boundaries representing the most ambiguous formant structures were chosen (see Table 1).
(2) The average category boundaries were found by pooling the pre-test results from the Estonian subjects. Comparison of the average boundaries and the pre-test results of a Finnish subject showed that Estonian and Finnish category boundary areas are overlapping well which allowed us to use the same ambiguous stimulus set for the Finnish group as well. However, in order to accommodate possible individual boundary variations, four different formant structures from averaged category boundary area in each vowel pair were selected for the main test (see stimuli numbers in Table 1). The formant values of prototype vowels and ambiguous stimuli in both sets are presented in Table 2 and Figure 1.

The vowel durations in both sets were manipulated from 60 to 140 ms in 20 ms steps, fundamental frequency was kept constant at 100 Hz. ABX and BAX series with an inter-stimulus interval of one second were constructed, where A and B were the prototype vowels of a vowel pair and X a stimulus of ambiguous quality. The duration of prototypes A and B were also varied from 60 to 140 ms in 20 ms steps, according to the duration of X, e.g. for X = 60 ms also A = B = 60 ms, etc.

The ABX/BAX setup was chosen to compensate for the potential influence of the preceding stimulus which occurred in a pilot test with single stimulus ordering type (see e.g. Repp, Crowder 1990 for a discussion of stimulus order effects in vowel perception).

3.2.2 Procedure

The main test was carried out with both Estonian and Finnish subject groups. For the Estonian group subject-specific stimuli set was used, whereas for the Finnish subjects the stimuli set based on the averaged category boundaries was presented. Testing conditions were the same as in the first experiment. The test was administered with Praat's multiple forced-choice test facility using balanced permutation for stimulus randomization. Each stimulus was repeated three times in both ABX and BAX contexts, adding up to a total of 450 stimuli (5 vowel pairs x 3 formant structures x 5 durations x 6 repetitions) for Estonian subjects and 600 stimuli (5 vowel pairs x 4 formant structures x 5 durations x 6 repetitions) for Finnish participants. Subjects had to answer (by clicking in one of two response boxes on the screen) the question "Does the sound you heard last resemble more the first or the second vowel?" A replay option was available during the listening test which allowed up to five repetitions of a stimulus. The experiment consisted of 5 blocks (each vowel pair in a separate block) with optional short breaks between blocks. On average, the test took about 45 minutes in the case of Estonian group and about 60 minutes in the case of Finnish group.

4. Results

The results of the main test confirm our hypothesis on the role of intrinsic duration in vowel category perception — the longer the duration of the ambiguous stimulus the more likely it is categorized as the more open vowel of a pair. Naturally, the results manifest some variation among languages, subjects and stimulus presentation order which are described next.

Perception scores of high vs low vowel in a pair for each formant setting and five duration steps were counted for each subject; a non-parametric statistic — the Kendall tau rank correlation coefficient — was calculated to evaluate the degree of correspondence between the vowel duration and perception score and to assess its significance.

Estonian subjects showed the best results in the expected direction in the case of the /i/-/e/ vowel pair: all 10 subjects had significant (on p-level < 0.1) correlation (tau = –0.63 to –0.95) at least for one formant setting; the results were slightly worse in the case of other high-mid vowel pairs: significant correlation was shown in the /y/-/ø/ pair by seven subjects

(tau = –0.67 to –0.95), two subjects showed almost no correlation; and in the /u/-/o/ pair by eight subjects (tau = –0.67 to –0.95).

In mid-low vowel pairs Estonian subjects showed different results — stimulus duration did not affect vowel category perception significantly. In the /e/-/æ/ pair only three subjects exhibited significant correlation in the expected direction (tau = –0.67 to –0.89), two subjects showed systematically significant correlation in the opposite direction (tau = 0.94 to 0.89; $p < 0.05$); in the /o/-/ɑ/ pair only two subjects showed significant correlation in expected direction (tau = –0.67 to –0.89), and three subjects showed significant correlation in the opposite direction (tau = 0.77 to 0.84), in mid-low vowel pairs five speakers had low correlation in both directions.

When processing the Finnish results it was found that in the case of one of four ambiguous stimuli (with the lowest or highest stimuli number depending on vowel pair) duration had almost no effect on vowel categorization (probably the formant structure was outside the boundary area for most Finnish subjects), therefore this formant structure was eliminated from the analysis. Consequently, we have valid perception data for three formant structures, as in the case of Estonian.

Unlike the Estonians, Finnish listeners demonstrated the expected results in all vowel pairs. Correlation was significant (at level $p < 0.1$) in the /i/-/e/ vowel pair for all ten subjects (tau = –0.63 to –0.95); in the /y/-/ø/ pair for six subjects (tau = –0.67 to –0.95), in the /u/-/o/ pair for eight subjects (tau = –0.77 to –0.95), in the /e/-/æ/ pair for eight subjects (tau = –0.67 to –0.89), and in the /o/-/ɑ/ pair for six subjects (tau = –0.67 to –0.84), two subjects did not perceive any category changes and two showed low correlations in both directions.

The stimulus presentation order (ABX vs. BAX) only had a minor random effect in some subjects and vowel pairs. This allowed us to pool the results of both conditions.

It can be seen that the use of two different stimulus sets for Estonian and Finnish groups did not cause any major discrepancies in the final results; different results in mid-low vowel pairs are certainly language-specific (this will be discussed in the next section) and not produced by different experimental settings.

After pooling the results in two subject groups over three formant structures and different stimulus presentation order, the averaged perception scores show high correlations between vowel category perception and stimulus duration in all vowel pairs for the Finnish subjects (/i/-/e/: tau = –1, $p = 0.017$; /y/-/ø/ and /u/-/o/: tau = –0.8, $p = 0.083$; /e/-/æ/ and /o/-/ɑ/: tau = –0.74, $p = 0.133$). The Estonian results are similar to the Finnish results only in high-mid vowel pairs: for /i/-/e/, /y/-/ø/ and /u/-/o/: tau = –1, $p = 0.017$; mid-low vowel pairs show totally different results: /e/-/æ/ tau = –0.316, $p = 0.32$, and /o/-/ɑ/: tau = 0.2, $p = 0.82$.

Linear regression plots and Pearson's regression coefficients for average perception scores are shown in the Figures 2—5.

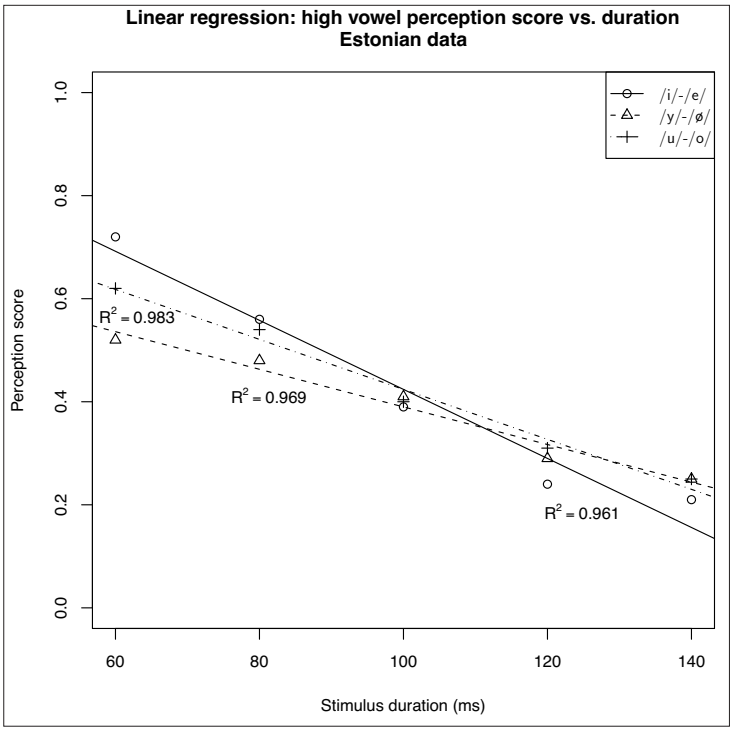

Figure 2. **Perception scores of high vowels in the high-mid vowel pairs vs. stimulus duration (average of all Estonian subjects).**

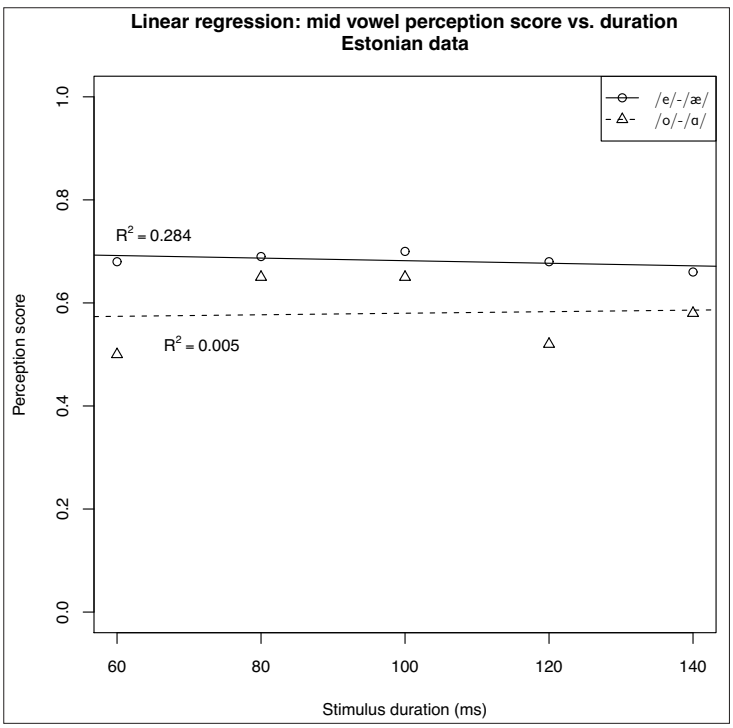

Figure 3. **Perception scores of mid vowels in the mid-low vowel pairs vs stimulus duration (average of all Estonian subjects).**

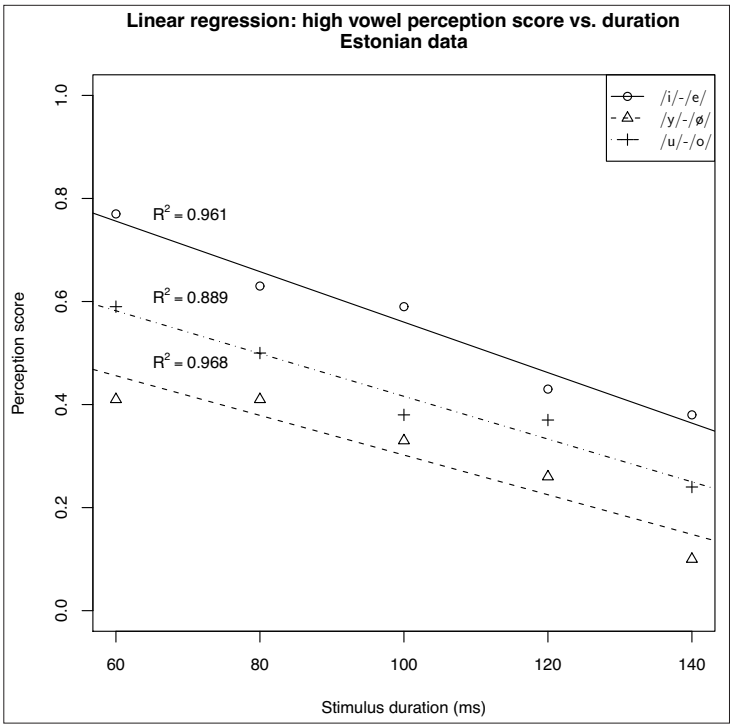

Figure 4. **Perception scores of high vowels in the high-mid vowel pairs vs. stimulus duration (average of all Finnish subjects).**

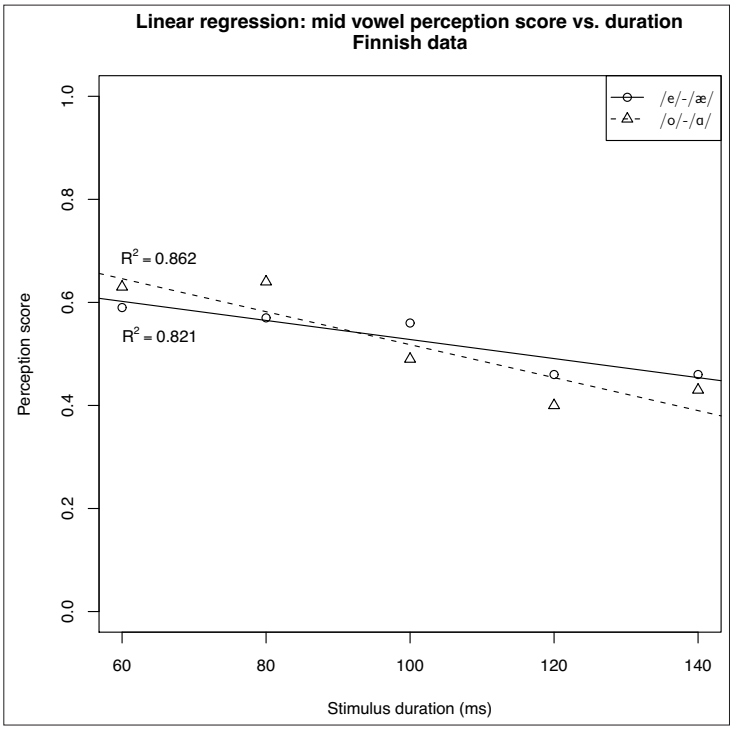

Figure 5. **Perception scores of mid vowels in the mid-low vowel pairs vs stimulus duration (average of all Finnish subjects).**

4. Discussion

The results show that duration of a vowel affects vowel category perception when spectral information is ambiguous, even in quantity languages. According to our tentative hypothesis this influence is based on differences between the intrinsic durations of high, mid and low vowels which are physiologically-based and thus outside the scope of speaker control. This hypothesis seems natural in the case of Estonian and Finnish where duration is intentionally controlled by a speaker on higher (phonemic and syllabic/foot) levels in order to produce contrastive quantity oppositions. However, different results for high-mid versus mid-low vowel pairs in the case of Estonian subjects cannot be explained by this hypothesis alone (assuming that the impact of intrinsic duration remains constant along the open-close dimension).

We should also consider the fact that our stimulus set includes a range of vowels from phonologically short to phonologically long types in both languages: in Estonian the category boundary lies around 90—120 ms depending on speech rate (Eek, Meister 2003); in Finnish the short/long (/a/ vs. /aa/) duration ratio is 1 : 2.3 with an average duration of 80 ms for /a/ and 180 ms for /aa/ (Isei-Jaakkola 2004), and the short/long category boundary for the first syllable of CV(V)CV(V) words for example is 116 ms (Ylinen, Shestakova, Alku, Huotilainen 2005). Our results cannot be explained by the category switch since there are no major quality differences between short and long vowels in either language (Eek, Meister 1998; Wiik 1965). If the short/long category change had an effect on quality perception, then the recognition scores would not fit the linear regression lines, instead a step-like change of the recognition score around the stimuli durations of 100 and 120 ms should be expected.

Another possible explanation is that intrinsic duration of vowels, attributed to physiological properties of the human vocal tract, is actually consciously controlled by the speaker also in quantity languages. This claim is supported by findings of studies on non-quantity languages like Russian (Kouznetsov 2001) and Spanish (Solé 2007) where intrinsic features of speech are found to be under control of a speaker to increase phonetic contrast between different phonological categories.

Yet, we would like to keep our initial hypothesis based on the physiological grounding of intrinsic features and to provide a different explanation for our results. According to theories modelling sound system evolution in different languages (Liljencrants, Lindblom 1972; Schwartz, Boë, Vallée, Abry 1997) vowels tend to maximize perceptual distances in the perceptual plane F1 vs. F'2 (F'2 is the so-called "perceptual second formant" integrating F2, F3 and F4). Depending on the number of vowels in a language, secondary articulatory contrasts (e.g. nasality, length, pharyngealisation) may be combined with (primary) quality contrasts in order to achieve the goal of maximal perceptual contrast (Schwartz, Boë, Abry 2007).

The Estonian vowel system includes nine vowel phonemes, divided into groups according to tongue height as: /i/, /y/, /u/ high, /e/, /ø/, /ɤ/, /o/ mid, and /æ/, /ɑ/ low vowels (the mid-vowel /ɤ/ is actually a non-low vowel as it extends to the high vowel area, too) (Eek, Meister

Table 3

Formant values of Estonian and Finnish vowel types

	Estonian vowel formants				Finnish vowel formants			
	F1, Hz	**F'2, Hz**	**F1, Bark**	**F'2, Bark**	**F1, Hz**	**F'2, Hz**	**F1, Bark**	**F'2, Bark**
/i/	259	3005	2.55	15.70	340	3145	3.43	15.99
/e/	419	2285	3.59	13.90	500	2771	4.92	15.17
/æ/	735	1366	6.23	10.48	675	2122	6.34	13.41
/y/	251	1858	2.55	12.52	340	2419	3.43	14.28
/ø/	428	1603	3.79	11.53	510	1936	5.01	12.79
/u/	275	549	2.76	5.34	400	780	4.01	7.10
/o/	434	630	3.98	5.99	535	985	5.22	8.44
/ɑ/	622	961	5.83	8.25	710	1355	6.60	10.43

Table 4

Perceptual distances (in Bark) of Estonian and Finnish vowel prototypes

	Perceptual distance, Bark	
Vowel pair	**Estonian**	**Finnish**
/i/-/e/	2.1	1.7
/y/-/ø/	1.6	2.2
/u/-/o/	1.4	1.8
/e/-/æ/	4.3	2.3
/o/-/ɑ/	2.9	2.4

1994). The Finnish eight-vowel system lacks the non-low back vowel /ɤ/, otherwise the division is similar to the Estonian one.

To evaluate perceptual distances between vowels the formant values F1—F4 measured in the acoustic analysis will serve as the basis: Estonian formant data is taken from Eek, Meister 1994 (Table 1), Finnish data is from Wiik 1965. The values of perceptual second formant F'2 have been calculated on the bases of F1—F4 data applying the Bladon-Fant formula (Bladon, Fant 1978). Perceptual distance between vowel types is calculated as a Euclidean distance using formant data F1 and F'2 from Table 3; the results are given in Table 4.

Comparing the perceptual distances between vowels in different vowel pairs we can see that in Estonian vowels the distance of high-mid pairs is shorter than that of mid-low pairs: in the case of front vowels 1.6—2.1 Bark for /y/-/ø/ and /i/-/e/ vs. 4.3 Bark for /e/-/æ/, and in back vowels correspondingly 1.4 Bark (/u/-/o/) vs. 2.9 Bark (/o/-/ɑ/).

Perceptual distances of Finnish vowels in high-mid vs. mid-low areas are almost equal as Finnish mid-vowels lie mainly in between high and low vowels: cf. 1.7—2.2 Bark for high-mid vowels and 2.3—2.4 Bark for mid-low vowels. This constitutes a possible explanation for the different results from Estonian and Finnish subjects.

We argue that in case of shorter perceptual distances intrinsic duration plays the role of a secondary feature to enhance perceptual contrast between vowels. When the perceptual distance is large enough (like in Estonian mid-low vowel pairs), the necessary perceptual contrast is guaranteed by spectral features alone and an additional duration cue is not needed.

5. Conclusions

Vowel duration in quantity languages is typically considered unrelated to vowel quality since duration is exploited to realise the quantity oppositions. Microprosodic universals, found also in Estonian and Finnish, however, suggest that durational variations might affect the perception of vowel quality also in languages with phonemic quantity oppositions. Our main experiment was designed to test this hypothesis by systematically varying vowel duration of stimuli with ambiguous vowel quality and recording native subjects' vowel identification decisions.

We were able to demonstrate a consistent correlation between vowel duration and perceived vowel quality both in Estonian and Finnish speakers. The results for both languages show that in the case of high-mid vowel pairs vowel openness correlates positively with stimulus duration; in mid-low vowel pairs such correlation was only found for some of the Finnish subjects. We put forward a tentative explanation for the difference between high-mid and mid-low pairs in the Estonian results: the difference between high-mid and mid-low pairs in Estonian results can be due to perceptual distance — in case of shorter distances between vowels prototypes, i.e. in the high-mid area of vowel space, intrinsic duration plays the role of a secondary feature to enhance perceptual contrast between vowels, whereas in the area of mid-low oppositions the distance is large enough to guarantee the necessary perceptual contrast by spectral features alone and an additional cue like intrinsic duration of a vowel is less essential.

Acknowledgements

All subjects having participated in the perception tests are gratefully acknowledged. Special thanks go to the anonymous reviewer whose well-founded critical observations we have tried to take into account as far as possible in our final version.

This research was partially supported by the target-financed theme No.0322709s06 of the Estonian Ministry of Education and Research and by the National Programme for Estonian Language Technology. The authors' cooperation has been supported both by the Academy of Finland and the Estonian Academy of Sciences.

Addresses:

Einar Meister
Tallinn University of Technology
E-mail: einar@ioc.ee
Phone: +372 6204200

Stefan Werner
University of Joensuu
E-mail: stefan.werner@joensuu.fi
Phone: +358 (0)13 251 4334

REFERENCES

Antoniadis, Z., Strube, H. W. 1984, Untersuchungen zur spezifischen Dauer deutscher Vokale. — Phonetica 41 (2), 72—87.

Black, J. W. 1949, Natural Frequency, Duration, and Intensity of Vowels in Reading. — Journal of Speech and Hearing Disorders 14, 216—221.

Bladon, A., Fant, G. 1978, A Two-Formant Model and the Cardinal Vowels. — Speech Transmission Laboratory Quarterly Progress and Status Report 19(1), 1—8.

Boersma, P., Weenink, D. 2008, Praat [Computer program] (Version 5.0.43). http://www.praat.org/.

Di Cristo, A. 1978, De la microprosodie à l'intonosyntaxe, Aix-en-Provence.

Eek, A. 1994, Studies on Quantity and Stress in Estonian, Tartu (Dissertationes Philologiae Estonicae Universitatis Tartuensis 4).

Eek, A., Meister, E. 1994, Acoustics and Perception of Estonian Vowel Types. — Experiments in Speech Processes, Stockholm (PERILUS. Phonetic Experimental Research at the Institute of Linguistics University of Stockholm 18), 55—90.

—— 1998, Quality of Standard Estonian Vowels in Stressed and Unstressed Syllables of the Feet in Three Distinctive Quantity Degrees. — LU XXXIV, 226—233.

—— 1999, Estonian Speech in the BABEL Multi-Language Database. Phonetic-Phonological Problems Revealed in the Text Corpus. — Proceedings of LP'98L. Item Order in Language and Speech, Prague, 529—546.

—— 2003, Domain of the Estonian Quantity Degrees. Evidence from Words Containing Diphthongs. — Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, 2039—2042.

—— Foneetilisi katseid ja arutlusi kvntiteedi alalt (I). Häälikukestusi muutvad kontekstid ja välde. — KK, 815—837.

Eerola, O., Savela, J., Laaksonen, J.-P., Aaltonen, O. 2002, Keston vaikutus suomen [y] / [i] -vokaalijatkumon kategorisointiin ja [i]-vokaalin prototyypin havaitsemiseen. — Fonetiikan Päivät 2002, Espoo, 115—122.

Isei-Jaakkola, T. 2004, Lexical Quantity in Japanese and Finnish, Helsinki.

Klatt, D. 1980, Software for a Cascade/Parallel Formant Synthesizer. — Journal of the Acoustical Society of America 67(3), 971—995.

Kouznetsov, V. 2001, Perceptual Role of Inherent Vowel Duration as Distinctive Feature in Russian. — Proceedings of the XI Session of the Russian Acoustical Society, Moscow, November 19—23, 2001, Moscow, 443—447.

Kouznetsov, V., Ott, A. 1987, Spectral Properties of Russian Stressed Vowels in the Context of Palatalized and Nonpalatalized Consonants. — Proceedings of the XIth International Congress of Phonetic Sciences 3, Tallinn, 117—120.

Ladefoged, P. 1964, A Phonetic Study of West African Languages. An Auditory-Instrumental Survey, Cambridge.

Lehiste, I. 1989, Current Debates Concerning Estonian Quantity. — Proceedings of the Sixth Annual Meeting of the Fenno-Ugric Studies Association of Canada, Lanham—New York—London, 77—86.

Liljencrants, J., Lindblom, B. 1972, Numerical Simulation of Vowel Quality Systems. The Role of Perceptual Contrast. — Language 48, 839—862.

McMurray, B. (in preparation), KlattWorks. A [Somewhat] New Systematic Approach to Formant-Based Speech Synthesis for Empirical Research [Computer program]. http://www.psychology.uiowa.edu/faculty/mcmurray/KlattWorks/default.html.

Meister, E., Meister, L., Werner, S. 2008, Microdurational Variations and Perception of Vowel Quality. — Fonetiikan päivät 2008, Tampere, 157—165.

Meister, E., Werner, S. 2006, Intrinsic Microprosodic Variations in Estonian and Finnish. Acoustic Analysis. — Fonetiikan Päivät 2006, Helsinki, 103—112.

Neweklowsky, G. 1975, Specific Duration and Specific Musical Pitch of Vowels. — Phonetica 32(1), 38.

Peterson, G. E., Lehiste, I. 1960, Duration of Syllable Nuclei in English. — Journal of the Acoustical Society of America 32(6), 693—703.

Repp, B. H., Crowder, R. G. 1990, Stimulus Order Effects in Vowel Discrimination. — Haskins Laboratories Status Report on Speech Research 103/104, New Haven, 111—124.

Rossi, M. 1972, Le seuil différentiel de durée. — Papers in Linguistics and Phonetics to the Memory of Pierre Delattre, The Hague, 435—450.

Schwartz, J., Boë, L., Abry, C. 2007, Linking Dispersion-Focalization Theory and the Maximum Utilization of the Available Distinctive Feature Principle in a Perception-for-Action-Control Theory. — Experimental Approaches to Phonology, Oxford, 104—124.

Schwartz, J., Boë, L., Vallée, N., Abry, C. 1997, The Dispersion-Focalization Theory of Vowel Systems. — Journal of Phonetics 25, 255—286.

Solé, M. 2007, Controlled and Mechanical Properties in Speech. A Review of the Literature. — Experimental Approaches to Phonology, Oxford, 302—321.

Solé, M., Ohala, J. J. [to appear], What Is and What Is not under the Control of the Speaker. Intrinsic Vowel Duration. — 10th Conference on Laboratory Phonology. http://aune.lpl.univ-aix.fr/~labphon10/abstracts/LP10_abstract_Sole.pdf

Van den Heuvel, H., Rietveld, T., Cranen, B. 1994, Methodological Aspects of Segment- and Speaker-Related Variability. A Study of Segmental Durations in Dutch. — Journal of Phonetics 22(4), 389.

Werner, S., Meister, E. 2008, Microdurational Influences on Perceived Vowel Quality. — Proceedings of the 3rd Baltic Conference on Human Language Technologies, Kaunas, October 4—5, 2007, Vilnius, 335—342.

Wiik, K. 1965, Finnish and English Vowels, Turku (Turun yliopiston julkaisuja B: 94).

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.

ЭЙНАР МЕЙСТЕР (Таллинн), *СТЕФАН ВЕРНЕР* (Йоэнсуу)

ДЛИТЕЛЬНОСТЬ ВЛИЯЕТ НА ВОСПРИЯТИЕ ГЛАСНЫХ В ЭСТОНСКОМ И ФИНСКОМ ЯЗЫКАХ

Идентификацию гласных в квантитативных языках принято считать независимой от длительности гласного, поскольку длительность как основной признак фонологических оппозиций не может быть определяющим признаком в других оппозициях. В статье исследуется роль микро-вариации длительности в восприятии гласного в эстонском и финском языках с помощью синтетических стимулов, содержащих фонологические оппозиции между гласными высокого и среднего подъема /i/-/e/, /y/-/ø/, /u/-/o/, а также среднего и низкого подъема /e/-/æ/, /o/-/ɑ/. Длительность стимулов варьирует от 60 до 140 мсек с шагом 20 мсек. В эксперименте участвовали 10 эстонских и 10 финских испытуемых.

Результаты эксперимента показывают позитивную корреляцию между открытостью гласного и его длительностью при стимулах, находящихся на границе между гласными высокого и среднего подъема, в обоих языках. При стимулах вблизи границы среднего и низкого подъема такой же результат показали только финские испытуемые, у эстонских испытуемых существенную корреляцию между открытостью гласного и его длительностью обнаружить не удалось. Мы выдвигаем гипотезу: микровариации длительности (собственная длительность гласного) выступают в роли секундарного признака только в том случае, когда перцептивное расстояние между акустическими целями гласных не достаточно для их дискриминации (т. е. в окружении гласных высокого и среднего подъема), но при достаточном расстоянии акустических целей (т. е. в окружении гласных среднего и низкого подъема) идентификация типа гласных осуществляется только при помощи спектральной информации.