 doi:10.3176/lu.2012.3.03

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

MICRODURATION IN FINNISH AND ESTONIAN VOWELS REVISITED: METHODOLOGICAL MUSINGS

Abstract. The influence of vowel duration on the perception of different vowel qualities in Finnish and Estonian has been the topic of several of our recent studies; for the present paper, we reconsider some of our methodological choices, comparing various different solutions. In the area of test design, timed group tests are evaluated as an alternative to our original self-paced individual test setup and test reliability is explored through repeated tests on the same subject. In the area of test evaluation, reaction time is added as a second dependent variable and a more sophisticated statistical evaluation is applied. All these methodological variations confirm the trends already visible in the results of our earlier studies.

Keywords: Estonian, Finnish, vowels, intrinsic duration.

1. Introduction

The studies on microprosody in several languages have established systematic differences in the intrinsic features of vowels — open vowels tend to have lower F0, higher intensity and longer duration than close vowels (e.g. Peterson, Lehiste 1960; Solé 2007; Di Cristo 1978; Wahlen, Levitt 1995; Meister, Werner 2006). Our recent studies address the intrinsic vowel duration in quantity languages like Estonian and Finnish, mainly focusing on the role of intrinsic duration on the perception of different phonological categories, i.e. vowel contrasts in close-open dimension and short vs. long durational oppositions. We have shown experimentally that in boundary conditions when spectral as primary features do not provide sufficient information for category discrimination in close-open vowel pairs, the intrinsic duration of vowels acts as a secondary feature facilitating the perceptual decision (Meister, Werner 2009). In a subsequent study we have found further evidence for the impact of intrinsic vowel duration by examining the categorical short vs. long distinction — the vowel quality (hence intrinsic duration of a vowel) plays a significant role in the discrimination of Estonian short vs. long phonological category (Meister, Werner, Meister 2011). The latter result is rather surprising since in quantity language like Estonian duration has to be intentionally controlled by a speaker to signal quantity contrasts and this ”higher order” control can ”override” the intrinsic features.

The aim of our current paper is to verify our previous findings on short vs. long category discrimination by different groups of subjects involving Estonian and Finnish subjects, and to address a number of methodological issues like different test setups, intra-subject variations in repeated experiments, different methods applied in the statistical analysis of the results.

2. Methods and data

2.1. Stimulus corpus

For the perception experiments a stimulus corpus involving short vs. long category oppositions in close vowel /i/ and open vowel /a/ in CV(:)CV carrier words was designed. The stimuli were created from the nonsense words /kaka/, /kiki/, /papa/, /pipi/, /tata/, and /titi/ pronounced in isolation by a native Estonian male speaker. In all words the duration of the stressed vowel (V1) was manipulated from 100 ms to 190 ms in 10 ms steps which consequently resulted in six stimulus sets from CVCV to CV:CV — /kaka/ vs. /ka:ka/, /papa/ vs. /pa:pa/, /tata/ vs. /ta:ta/, /kiki/ vs. /ki:ki/, /pipi/ vs. /pi:pi/, /titi/ vs. /ti:ti/. The durations of the other segments were kept constant (C1(burst) = 25 ms for /k/, 15 ms for /p/ and /t/; C2 = 75 ms; V2 = 240 ms); the F0 was set to a constant value of 100 Hz in both vowels. The number of different stimuli in all sets was 10. The manipulation of stimuli was done with Praat (Boersma, Weenink 2011).

In Estonian, the stimulus sets constitute a continuum from a word in quantity one (Q1) to a word in quantity two (Q2) achieved by changing the duration of the first-syllable vowel.

2.2. Test variations

Factors whose potential influence we wanted to assess were:

- test setup: self-paced (individual) vs. timed (group) test
- test-retest intra-subject variation
- test evaluation: reaction times as additional support
- test evaluation: statistical modelling

To investigate the possible effect of imposing a time limit on the subjects we designed two slightly different group versions of our quantity perception tests, one of which was administered to Estonian subjects, the other to Finnish subjects. The Estonian version contained the full set of different stimuli from the individual tests – two vowel qualities and three consonant articulation places — with each stimulus played three times and an inter-stimulus interval of five seconds, whereas the Finnish version only used three of the six stimuli (/kaka/, /kiki/, /papa/) but played every stimulus five times using an inter-stimulus interval of three seconds. The Estonian group involved 40 subjects whereas 30 subjects where native speakers of Estonian (EST-L1) and 10 non-native subjects with Russian-language background (EST-L2); the Finnish group involved 17 native speakers (FIN-L1).

For both native groups short vs. long category discrimination is natural since both Estonian and Finnish exploit the duration cue contrastively; also L2 subjects are able to discriminate Estonian short and long contrasts despite non-categorical role of duration in Russian (Meister, Meister 2011).

In order to check for test-retest variation, one native Finnish and two native Estonian subjects underwent the same test several times; for the Finnish subject, reaction times were now also recorded. Instead of a linear regression analysis of response frequency in terms of duration (as in Meister, Werner 2009), we fitted more complete binomial logistic regression models, with and without random effects.

3. Results

3.1. Short vs. long boundaries

Overall group test results are in line with our previous studies: vowel openness correlates positively with stimulus duration in all subjects' groups (Figure 1). In EST-L1 group the boundary mean in the case of high vowel lies at 146.8 ms and

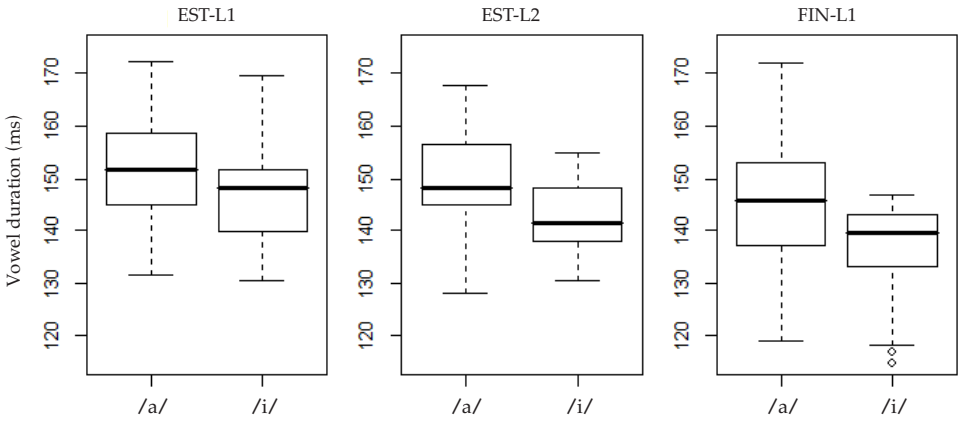

Figure 1. **Boxplots of the distributions of Estonian and Finnish speakers' category boundaries between "long" and "short" in low and high vowels.**

in the case of low vowel 151.7 ms, in EST-L2 the categorical boundary values are slightly lower — 142 ms and 147.3 ms for high and low vowel, correspondingly. The boundary values for the Finnish group lie even at shorter vowel durations — at 135.9 ms in the case of high vowel and at 144.5 ms in the case of low vowel. The difference between the two means is significant in two native groups, in EST-L1 group at the 0.001 level (Welch two-sided t-test, $t = 3.9$; $df = 174$; $p < 0.001$) and in FIN-L1 group at the 0.01 level ($t = 2.7$; $df = 35$; $p < 0.01$); in the EST-L2 group the difference in category mean values between low and high vowel turned out to be insignificant ($t = 1.4$; $df = 35.9$; $p = 0.17$).

The variation of responses broken down by stimulus duration is shown in Figure 2 (the difference in frequency range is due to the higher number of observations per duration and subject in the Estonian test). The area of indecision around the category boundary seems to spread out slightly more in the Estonian data. But this can be due to, at least partly, the greater variation of segmental contexts in the Estonian stimulus material and the larger number of Estonian test subjects (here EST-L1 and EST-L2 groups are pooled together).

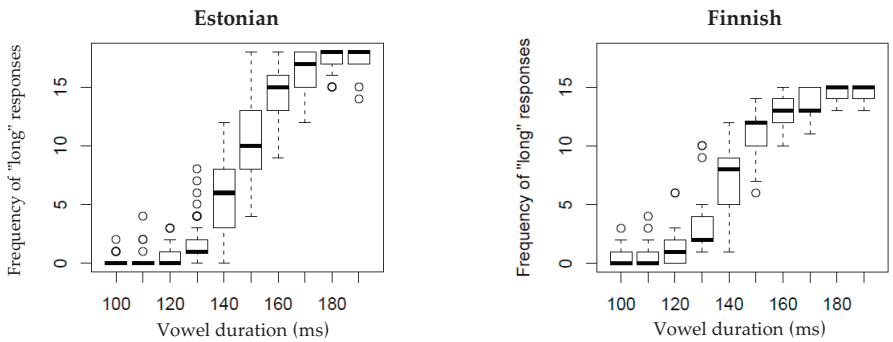

Figure 2. **Boxplots of the proportions of Estonian and Finnish speakers' "long" responses across stimulus durations. The whiskers extend to 1.5 times the interquartile range, indicating a 95% confidence interval for the difference in medians.**

3.2. Test setup

Test setup does not seem to have a systematic influence on the perception test results in our one-subject Finnish case study. The subject's category boundary ranged

from 148.4 ms to 133.7 ms in the four identical individual tests which were self-paced, and was 142.3 ms in the time-controlled group test.

3.3. Test-retest variation

As illustrated in the mosaic plots of Figures 3 and 4 intra-subject test-retest variation turns out to be moderately high for our Finnish case who went through the test six time (see previous section), but minimal for the two Estonian two-test cases: cate-

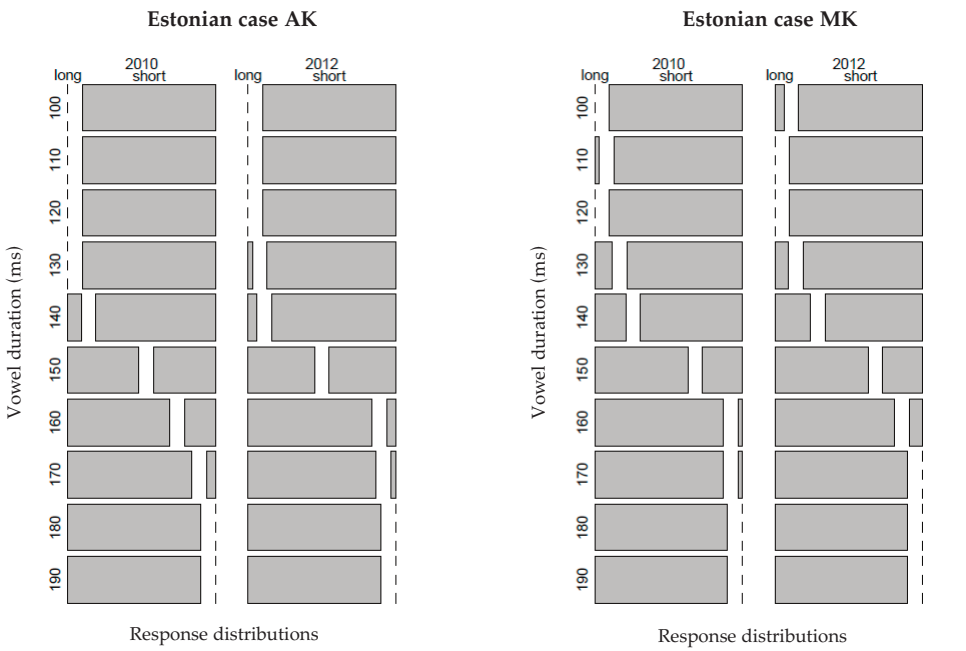

Figure 3. **Test-retest comparison of response distributions for two Estonian subjects.**

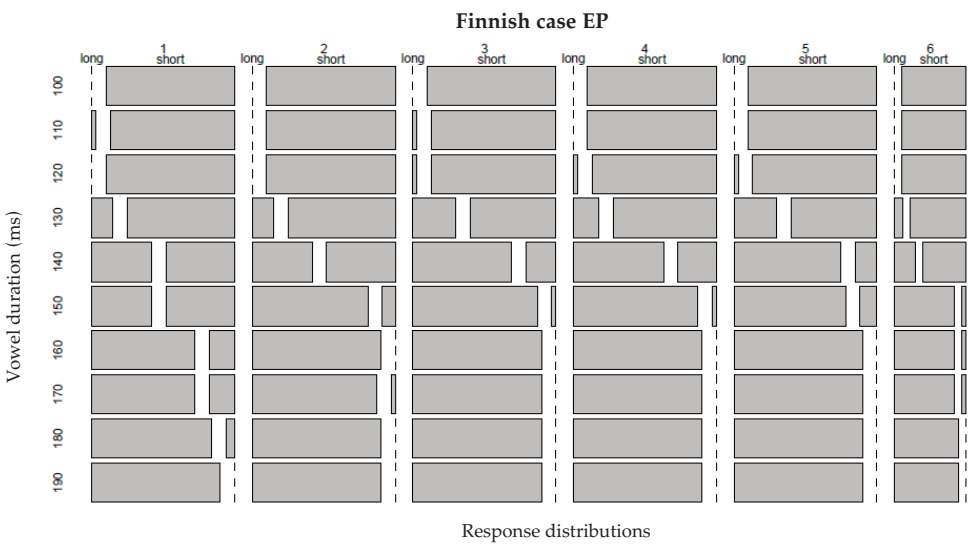

Figure 4. **Test-retest comparison of response distributions for Finnish subject.**

gory boundaries are at 151.7 and 150.0 for subject AK and at 144.7 in both tests for subject MK. All in all, the category boundaries between long and short are not affected in a way that would challenge our overall results for both language and intra-subject variation between tests does not exceed within-test intra-subject variation for repeated stimuli. Even in the Finnish case, median durations for long vs. short only fluctuate between adjacent conditions: 170 vs. 160 ms and 120 vs. 110 ms for long and short responses, respectively.

3.4. Reaction time

In two of the four self-paced tests of Finnish subject EP reaction times were measured. As can be seen from Figure 5, there seems to be a slight trend for reaction time to increase towards the category boundary which lies at 135.5 ms for these two tests. There is a weak but significant negative correlation ($r = –0.16$, $p < 0.001$) between the squared distance of stimuli's' duration from the category boundary and reaction time. If a similar trend could be observed in other subjects as well it would lend additional support to our estimation of the category boundaries.

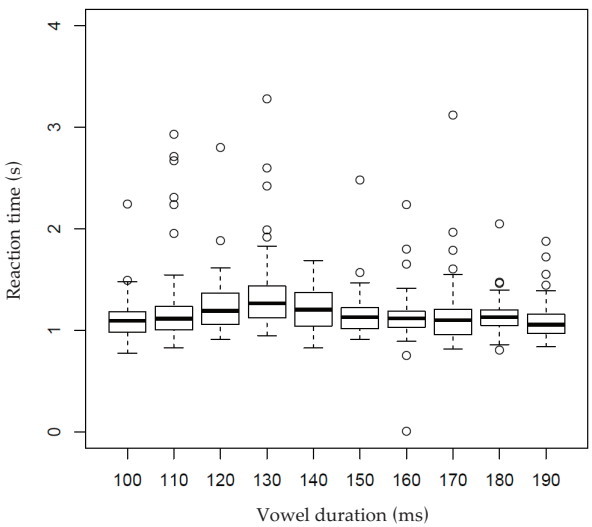

Figure 5. **Boxplots of Finnish subject EP's reaction times (from two tests). The whiskers extend to 1.5 times the interquartile range, indicating a 95% confidence interval for the difference in medians (the zero value at 160 ms must be due to an inadvertent keypress).**

3.5. Statistical models

We fitted binomial logistic regression models using R's glm() and glmer() functions. The mixed model analyses adding subject, stimulus, and/or presentation order as random effects did not produce results that significantly differed from the fixed-effects-only model: the only relevant factor, in addition to stimulus duration, is the consonantal context. Vowel openness, although affecting the categorical boundary in a minimal model with duration as the only factor (146 ms vs. 151 ms and 136 ms vs. 144 ms for high vs. low in Estonian and Finnish speakers, respectively), does not improve the model fit significantly. Subjects' sex and age reduced model deviance even less. Table 1 shows as an example the deviance analysis of a three-factor model for the Finnish group data.

Table 1

Analysis of deviance table for a binomial logistic regression model (logit link function) of duration perception in the Finnish group test with factors duration, consonant place of articulation and vowel openness

	Df	Deviance	Resid. Df	Resid. Dev.	Pr(Chi)
NULL			2549	3523.4	
dur	1	1733.08	2548	1790.4	2e-16 ***
cons	1	108.41	2547	1682.0	2e-16 ***
vow	1	1.47	2546	1680.5	0.2247

4. Discussion

Our new tests with Estonian and Finnish subjects lend further support to our previous findings on the connection between intrinsic vowel duration and perceptual vowel categorization in quantity languages. Our case study of reaction time measurements in addition to perceptual ratings also shows the same trend.

On the basis of our one-subject case study it seems that the influence of variations in the test set-up can be neglected but more data will be needed to prove this point. Finally, more sophisticated statistical analyses with mixed models instead of fixed-factors-only models do not introduce new insights into our data.

All in all, the collection of results from new data and reconsiderations of methodological solutions presented here consolidates the concept of micro- and macroduration interplay developed already in our earlier studies.

Acknowledgement

This work has been partly supported by the target-financed theme No. 0140007s12 of the Estonian Ministry of Education and Research. We are grateful to Lya Meister for conducting the perception experiments with Estonian listeners and to all subjects who participated in the study.

Addresses

Stefan Werner
Department of Linguistics, University of Eastern Finland
E-mail: stefan.werner@uef.fi

Einar Meister
Institute of Cybernetics, Tallinn University of Technology, Estonia
E-mail: einar@ioc.ee

REFERENCES

Boersma, P., Weenink, D. 2011, Praat: doing phonetics by computer (Version 5.2.09) [Computer program]. http://www.praat.org/.

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

Meister, L., Meister, E. 2011, Perception of the Short vs. Long Phonological Category in Estonian by Native and Non-Native Listeners. — Journal of Phonetics 39, 212—224.

Meister, E., Werner, S. 2006, Intrinsic Microprosodic Variations in Estonian and Finnish. Acoustic Analysis. — Fonetiikan Päivät 2006, Helsinki (Publications of the Department of Speech Sciences, University of Helsinki 53), 103—112.

—— 2009, Duration Affects Vowel Perception in Estonian and Finnish. — LU XLV, 161—177.

M e i s t e r, E., W e r n e r, S., M e i s t e r, L. 2011, Short vs. Long Category Perception Affected by Vowel Quality. — ICPhS XVII. The 17th International Congress of Phonetic Sciences, Hong Kong, 17—21 August 2011, Hong Kong, 1362—1365.

P e t e r s o n, G. E., L e h i s t e, I. 1960, Duration of Syllable Nuclei in English. — Journal of the Acoustical Society of America 32, 693—703.

S o l é, M. J. 2007, Controlled and Mechanical Properties in Speech. A Review of the Literature. — Experimental Approaches to Phonology, Oxford, 302—321.

W a h l e n, D. H., L e v i t t, A. G. 1995, The Universality of Intrinsic F0 of Vowels. — Journal of Phonetics 23, 349—366.

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

О МИКРОДЛИТЕЛЬНОСТИ ФИНСКИХ И ЭСТОНСКИХ ГЛАСНЫХ. МЕТОДОЛОГИЧЕСКИЕ РАЗМЫШЛЕНИЯ

Влияние длительности гласного на восприятие качества гласных в финском и эстонском языках изучено в нескольких наших недавних исследованиях, в данной статье пересмотриваются некоторые из наших методологических подходов. В части разработки слуховых экспериментов групповой тест сравнивается с исходным индивидуальным, а достоверность результатов проверяется повторными экспериментами с одним субъектом. Для дополнительной оценки теста измеряется время реакции и при анализе результатов используются более сложные статистические методы. Все эти методологические вариации подтверждают результаты наших предыдущих исследований.