

The production and perception of Estonian quantity degrees by native and non-native speakers

Lya Meister, Einar Meister

Laboratory of Phonetics and Speech Technology
Institute of Cybernetics at Tallinn University of Technology, Estonia

lya@phon.ioc.ee, einar@ioc.ee

Abstract

The paper studies the production and perception of Estonian quantity oppositions by native Estonians (L1 subjects) and non-native speakers with Russian-language background (L2 subjects). Estonian quantity system involves three contrastive prosodic patterns referred to as short (Q1), long (Q2) and overlong (Q3) quantity degrees. These phonological contrasts are manifested by a complex interaction of durational and tonal cues in a disyllabic foot. For L2 subjects the Estonian quantity contrasts constitute a difficult issue in both perception and production since there are no similar patterns in Russian to rely on.

The production of quantity contrasts was examined in two-syllable CVCV-words read in sentence context; for the perception experiments a stimulus set involving isolated CVCV-words with varying duration of vowels in both syllables and varying F0 contour was created. Ten L1 and ten L2 subjects participated in both the reading of test sentences and the perception experiment.

The results showed that L2 subjects were successful in distinguishing Q1 and Q2 patterns in production and perception, but failed in distinguishing Q2 and Q3 patterns in both tasks.

Index Terms: L2 speech, quantity degrees, categorical perception, category boundary, Estonian, Russian

1. Introduction

Phonetic characteristics of the Estonian ternary quantity system and its phonological interpretations have been discussed in numerous papers (see e.g. [1], [2], [3], [4], etc.). The nowadays dominant view suggests that the domain of Estonian quantity contrasts is a disyllabic sequence consisting of the stressed and the following unstressed syllable (e.g. [5], [6], [7], [8]). Phonetically, the quantity oppositions are characterized by varying durational relations of the stressed and the unstressed syllable in a foot and signaled by complex tonal patterns. The quantity oppositions are manifested in vowels and diphthongs of the stressed syllable, e.g. Q1 *sada* /sata/ 'hundred', nom.sg.; Q2 *saada* /saata/ 'to send', sg.imperat.; Q3 *saada* /saa:ta/ 'to get'; Q2 *koera* /koera/ 'dog', gen.sg.; Q3 *koera* /koe:ra/ 'dog', part.sg., or in consonants and consonant clusters between the first and second syllable vowels, e.g. Q1 *kala* /kala/ 'fish', nom.sg.; Q2 *kalla* /kalla/ 'arum', nom.sg.; Q3 *kalla* /kal:la/ 'pour', 2.sg. imperat.; Q2 *lehma* /lehma/ 'cow', gen.sg.; Q3 *lehma* /leh:ma/ 'cow', part.sg. There is no vowel quantity contrast in the unstressed syllable; however, due to the tendency towards foot isochrony the duration of the second syllable vowel is inversely proportional to the duration of the first syllable vowel. A binary consonant contrast for stop consonants is possible also in the word-final position, but not in word-initial position.

Traditionally, the ratio of the first and the second syllable duration has been used as a primary cue to distinguish the quantity contrasts. As pooled from different studies, for Q1 the typical ratio is 0.6-0.8, for Q2 1.4-1.7, and for Q3 2.2-3.4 [9]. The tonal patterns of the foot as the second important cue, has the following typical patterns: in Q1 and Q2 F0 is rising during the stressed and falling in the unstressed syllable, in Q3 the F0 peak occurs earlier – it is located around the first half of the stressed vowel and falls during the rest of the stressed syllable to the end of the unstressed syllable [1], [2], [3].

For learners of Estonian as a foreign language the complexity of Estonian quantity system is difficult to acquire, especially for learners whose native language does not exploit duration cue contrastively. For example, Russian is one of these languages where duration of speech segments is not contrastive. Instead, duration in Russian acts as the main cue for word stress – duration of stressed-syllable vowels is longer than the duration of vowels in unstressed syllables [10].

The "Desensitization" Hypothesis [11] suggests that duration is an easily accessible acoustic cue independent of the role of duration in subjects' L1. Thus the contrastive variation of vowel duration in Estonian should be easily perceived by L2 subjects, regardless different roles of duration in Estonian and Russian.

The Feature Hypothesis [12] states that "L2 features not used to signal phonological contrasts in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner's production of the contrast based on this feature". According to this hypothesis L2 subjects with a Russian-language background should face difficulties in distinguishing phonological contrasts of Estonian in both perception and production.

The perception of Estonian short/long vowel categories by L1 and L2 subjects has been studied in three contexts: isolated vowels, CVC, and CVCV units [13]. The results demonstrated differences in binary (short vs. long) category perception by L1 and L2 subjects in all stimulus sets revealing different perception strategies of two groups. However, L2 subjects were quite successful in distinguishing the Estonian short vs. long categories despite the non-categorical use of the duration cue in their native language.

The current study extends the scope to the production and perception of more complex prosodic patterns, i.e. ternary quantity oppositions in Estonian.

2. Method

2.1. Subjects

In the study ten (5 male, 5 female) native speakers of Estonian and ten (5 male, 5 female) non-native subjects with Russian as their L1 were involved. L1 subjects (age 21-54, median 26.5) came from monolingual Estonian-speaking families and represent standard Estonian pronunciation.

The L2 subjects (age 21-33, median 24.5) were born in monolingual Russian speaking families living in the north-east of Estonia or in the capital area; they were educated in Russian at basic and high schools and have or are currently acquiring a university degree in Estonia. Most of the L2 subjects started to learn Estonian in school at the age of 6-13, one subject at the age of 16 and one at the age of 20. All L2 subjects use Estonian almost every day at university or at their place of work, but at home they communicate in Russian (except for one subject). None of the subjects reported any language impairment

2.2. Speech material

The Estonian Foreign Accent Corpus was used in the acoustic analysis. The corpus has been collected in a sound-proof room using high quality recording facilities. Each subject has read aloud a text corpus (140 sentences including the main phonological oppositions of Estonian, two short passages, and four prompts to elicit spontaneous speech). In the current study a part of the corpus – a subset of 27 sentences involving triplets of segmentally identical disyllabic target words in the quantities Q1, Q2 and Q3, representing the structures CVCV, CVVCV and CVV:CV (9 words in all structures) was used. The target words were segmented manually on the word and phone levels using Praat [14]; the duration of vowel segments was measured using a Praat script.

2.3. Stimulus sets and testing procedure

The CVCV stimulus corpus was created from a Q2 base word *saada* /saata/ 'to send', sg.imperat. pronounced by a native Estonian male speaker. The stimulus sets from Q1 to Q3 were generated by manipulating the duration of V1 from 80 ms to 160 ms in 20 ms steps and the duration of V2 from 60 to 140 ms in 20 ms steps (Figure 1); the duration of C1 and C2 was set to the constant value of 80 ms. Consequently, the first stimulus set, referred to as CVCV(1), represents the continuum from the lexical item in Q1 *sada* /sata/ 'hundred, nom.sg', to the lexical item in Q2 *saada*, and the second set, referred to as CVCV(2), from the lexical item in Q2 *saada* to the lexical item in Q3 *saada* /saa:ta/ 'to get'.

Two subsets of CVCV(1) and four subsets of CVCV(2) were created by manipulating the F0 contour as follows (all subsets include 10 different stimuli):

- 1a – F0 was kept constant at 130 Hz in both vowels;
- 1b – F0 was modeled to follow its typical course in natural Q1 and Q2 words, i.e. rising from 130 Hz to 160 Hz during the first vowel and falling from 130 Hz to 100 Hz during the second vowel;
- 2a – F0 was kept constant at 130 Hz in both vowels;
- 2b – F0 as in CVCV(1b);
- 2c – F0 was modeled as a typical contour in natural Q3 words, i.e. rising from 130 Hz to 160 Hz during the first third of V1 duration and then falling to 130 to the end of V1; the fall continues during V2 from 130 Hz to 100 Hz at the end of the second vowel;
- 2d – F0 was rising from 130 Hz to the peak value of 160 Hz and falling back to 130 Hz within the first vowel and the location of the peak was moved from the beginning of V1 to the end of V1 depending on the duration of V2.

The listening tests were conducted in a quiet room and stimuli were presented to subjects via high-quality headphones from a laptop computer. The test was administered with Praat's multiple forced-choice test facility. The listeners had to

decide on quantity category in a binary identification task by clicking on one of the two boxes labeled as Q1 and Q2 in the first two subsets and as Q2 and Q3 in the rest four subsets. In total 180 stimuli (10 stimuli x 3 repetitions x 6 subsets) were presented to each subject. The duration of the test was ca 20 minutes.

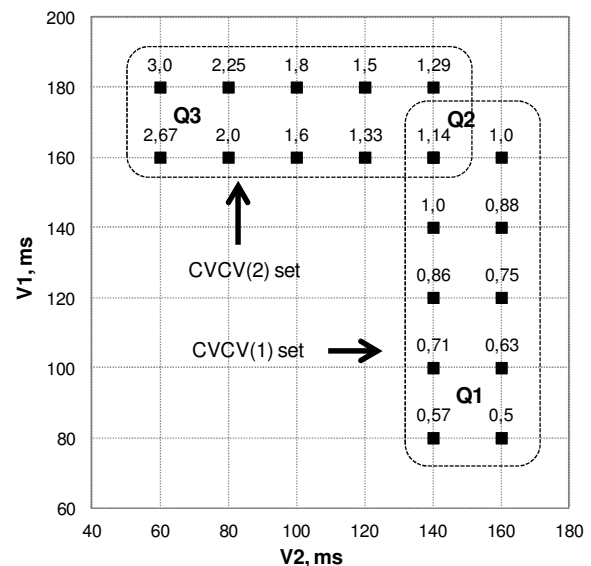


Figure 1. The stimuli (■) of two sets represented as the combinations of V1 and V2 duration; the numbers above stimuli correspond to V1/V2 duration ratios.

3. Results

3.1. Production of quantity degrees

For each speaker, 27 words (3 structures x 9 words) were analyzed, i.e. the durations of V1 and V2 were measured and the duration ratio V1/V2 was calculated for each word. The analysis of variance (ANOVA) with two grouping variables Language (L1 and L2) and Quantity (Q1, Q2 and Q3), and where applicable, *TukeyHSD* post-hoc tests were used.

In L1 speech both V1 and V2 participate in signaling the quantity oppositions resulting in three distinct V1/V2 duration ratios. Consequently, in L1 group Quantity had an effect on both V1 [F(2, 267) = 330.9; p < 0.001] and V2 duration [F(2, 267) = 113.1; p < 0.001], as well as on V1/V2 duration ratio [F(2, 267) = 523.1; p < 0.001]; the post-hoc tests showed significant differences in V1 and V2 durations in all word structures (p < 0.001).

In L2 group Quantity had an effect on V1 duration [F(2, 267) = 84; p < 0.001] and on V1/V2 ratio [F(2, 267) = 49.4; p < 0.001], but not on V2 duration [F(2, 267) = 1.66; p = 0.19]. The post-hoc tests revealed that V1 durations differed significantly in Q1 and Q2 (and Q3) (p < 0.001), but not in Q2 and Q3 (p = 0.99); in V1/V2 ratio the differences were significant in Q1 and Q2 (and Q3) (p < 0.001), but not in Q2 and Q3 words (p = 0.87).

Comparison of L1 and L2 groups showed that the groups had highly similar results in V1 and V2 durations (and consequently, V1/V2 ratio) in the case of Q2, but differed significantly in Q1 and Q3 (Table 1).

V1 and V2 durations of L1 subjects are highly similar to those reported in a numerous previous studies (e.g. [2], [5], [6], [8]) and exhibit the isochronic duration patterns of a foot – the duration of V2 is inversely proportional to the duration of V1 resulting in clearly distinct V1/V2 ratios for Q1, Q2 and Q3 (Figure 2, left). In L2 speech a clear contrast is produced

between Q1 and Q2, whereas the contrast is exhibited only by differences in V1 duration, but not in V2 duration. Q2 and Q3 are produced almost similarly; no difference in V1 and V2 durations is observed (Figure 2, right).

Table 1. L1 and L2 mean durations (in ms) and standard deviations of V1 and V2, and V1/V2 duration ratios in three quantity degrees.

	Quantity	L1		L2		L1-L2	
		Dur	Sd	Dur	Sd	F(1, 178)	p
V1	Q1	82	15.6	93	28.4	9.2	< 0.01
	Q2	144	29.0	148	33.1	0.9	0.4
	Q3	174	26.7	149	38.0	25.7	< 0.001
V2	Q1	112	28.3	97	40.1	10.4	< 0.01
	Q2	85	25.0	89	34.6	1.1	0.3
	Q3	61	12.2	88	31.3	56.1	< 0.001
V1/V2	Q1	0.8	0.03	1.1	0.52	32.4	< 0.001
	Q2	1.8	0.02	1.8	0.54	0.05	0.8
	Q3	2.9	0.01	1.8	0.63	136.4	< 0.001

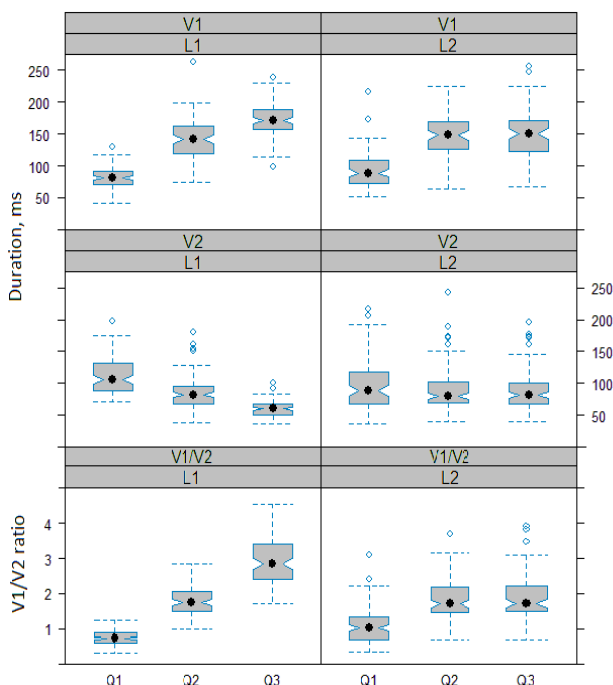


Figure 2. The box-plots demonstrating the distribution of V1 and V2 durations and V1/V2 ratio by Quantity in L1 and L2 group.

3.2. Perception of quantity degrees

The output of the perception experiment is a categorization function for each stimulus set that plots the score of one of the two alternative responses across the continuum. The categorization functions in two groups were obtained in two steps – first, the individual response data was interpolated with a *probit*-function and subjects' category boundaries were calculated as the 50% cross-over point of the fitted curve; second, the individual interpolated data was averaged over all subjects in a group. The individual category boundary values were analyzed using ANOVA.

L1 and L2 groups demonstrated similar results in stimulus sets involving Q1 vs. Q2 oppositions (Figure 3) – for both groups the duration of the vowel in the first (stressed) syllable was the most significant factor in Q1 vs. Q2 categorization

(L1: [F(4, 90) = 149.5, p < 0.001]; L2: [F(4, 90) = 106.4, p < 0.001]), the duration of the vowel in the second (unstressed) syllable turned out to be insignificant for both groups. The F0 contour was marginally significant for the L1 group only [F(1, 90) = 2.9; p < 0.1] resulting in a shorter boundary values of V1 (cf. stimulus sets 1a and 1b in Table 2).

In the Q2 vs. Q3 categorization task L1 and L2 differed significantly (Figure 4). For the L1 group the durations of V1 [F(1, 360) = 195.6, p < 0.001] and V2 [F(4, 360) = 112.6, p < 0.001] as well as the F0 contour [F(3, 360) = 30.9, p < 0.001] had significant effect, their interactions were significant, too. Q2-Q3 category switch due to change of V2 duration was only observed in those stimulus sets in which other factors (V1 duration and F0 contour) favored it. The duration of V2 and the duration ratio V1/V2 corresponding to the Q2-Q3 category boundary varied to a great extent depending on V1 duration and F0 contour (cf. stimulus sets 2a, b, c, d in Table 2). Thus, Q2 vs. Q3 categorization cannot be predicted by the duration ratio V1/V2 alone, also the F0 contour should be taken into account.

Table 2. L1 and L2 mean durations (in ms) of V1, V2 and V1/V2 duration ratios corresponding to category boundaries (in bold) in six stimulus sets.

Stimulus set	L1			L2			
	V1	V2	V1/V2	V1	V2	V1/V2	
Q1 vs. Q2	1a	120	140	0.9	122	140	0.9
		120	160	0.8	120	160	0.8
	1b	115	140	0.8	124	140	0.9
		117	160	0.7	120	160	0.8
Q2 vs. Q3	2a	160	73	2.2	160	65	2.4
		180	128	1.4	180	-*	-*
	2b	160	59	2.7	160	87	1.8
		180	92	2.0	180	-*	-*
	2c	160	95	1.7	160	128	1.3
		180	-*	-*	180	-*	-*
	2d	160	82	2.0	160	144	1.1
		180	110	1.6	180	-*	-*

* No category switch took place; all stimuli in a set were perceived as Q3.

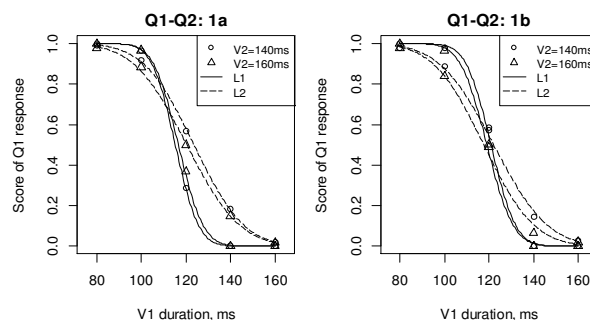


Figure 3. Average categorization curves of L1 and L2 group in stimulus sets involving Q1 vs. Q2 oppositions.

For the L2 group the main factor in Q2 vs. Q3 categorization was the duration of V1 [F(1, 360) = 78.4, p < 0.001] – stimuli with shorter V1 were classified as Q2 and stimuli with longer V1 as Q3; the duration of V2 and the F0 contour had no effect. The apparent Q2-Q3 category boundary observed for the L2 group cannot be a real boundary (since in their native language there is no such opposition available); the discovered boundary was rather induced by the design of the perception experiments and reflects the response strategy of the L2 subjects.

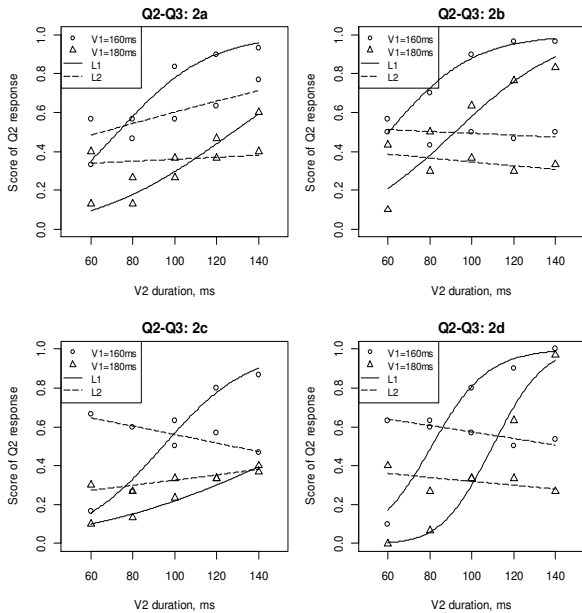


Figure 4. Average categorization curves of L1 and L2 group in stimulus sets involving Q2 vs. Q3 oppositions.

4. Discussion and conclusions

In order to compare production and perception of Estonian quantity degrees, the production data and the perceptual category boundaries found in the stimulus sets 1b and 2d (in these sets F0 contour follows the typical pattern of natural words and thus the comparison with production data is more adequate than in the case of other stimulus sets) have been plotted onto V1 vs. V2 space (Figure 5).

As expected, the results of the L1 group show good harmony between production and perception of quantity oppositions – in Q1, Q2 and Q3 words the temporal characteristics of vowels in the first and second syllable (in combination with an adequate F0 contour) result in three different patterns which are well separated from each other by the perceptual boundaries of Q1-Q2 and Q2-Q3 (Figure 5, left).

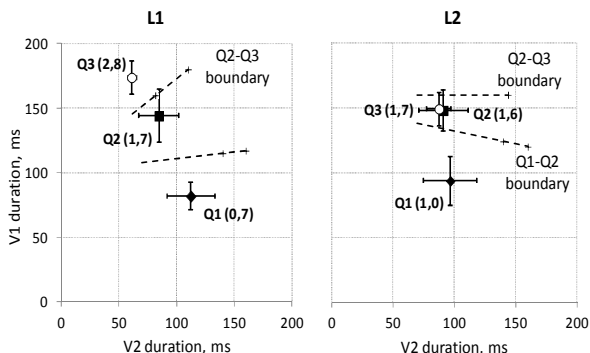


Figure 5. Average V1 and V2 durations and standard deviations in Q1(○), Q2(■) and Q3(◆) words and category boundaries (---) of L1 and L2 group.

In the L2 group the results show good correspondence between production and perceptual discrimination of Q1 and Q2 words, similar to the L1 group, but the results of production and perception of Q2 vs. Q3 contrasts are very different – L2 subjects are not able to perceptually

discriminate Estonian Q2 and Q3 structures nor to produce different patterns for Q2 and Q3 words in L2 speech (Figure 5, right).

The L2 results in Q1 vs. Q2 categorization support the desensitization hypothesis [11] – the variable V1 duration in Q1 and Q2 appeared to be salient and easily accessible cue for L2 subjects. However, the variable duration patterns of Q2 and Q3 turned out not to be easily accessible for the subjects with Russian-language background, and thus Q2 vs. Q3 results tend to support the feature hypothesis [12].

The strategy of L2 subjects to reduce the complex interaction of temporal and tonal cues in Q2 vs. Q3 contrast to a simple short vs. long contrast in V1 has failed. There is no phonological contrast available in Russian that would use duration and pitch cues similarly to the Estonian Q2 vs. Q3 contrast such that L2 subjects could rely on it in both perception and production.

5. Acknowledgements

This research was supported by the target-financed theme No. 0140007s12 of the Estonian Ministry of Education and Research and by the National Program for Estonian Language Technology.

6. References

- [1] Lehiste, I., "Segmental and syllabic quantity in Estonian", *American Studies in Uralic Linguistics 1*, Bloomington: Indiana University Press, 21-82, 1960.
- [2] Liiv, G., "Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid", *Keel ja Kirjandus*, 412-424, 480-490, 1961.
- [3] Eek, A., "Observations on the duration of some word structures: I", *Estonian Papers in Phonetics*, 18-31, 1974.
- [4] Lehiste, I. and Ross, J., "Estonian Prosody: Papers from a Symposium", Tallinn: Institute of Estonian Language, 1997.
- [5] Lehiste, I., "Search for Phonetic Correlates in Estonian Prosody", in I. Lehiste and J. Ross [Eds], *Estonian Prosody: Papers from a Symposium*, Tallinn: Institute of Estonian Language, 11-35, 1997.
- [6] Eek, A. and Meister, E., "Simple perception experiments on Estonian word prosody: foot structure vs. segmental quantity", in I. Lehiste and J. Ross [Eds], *Estonian Prosody: Papers from a Symposium*, Tallinn: Institute of Estonian Language, 71-99, 1997.
- [7] Krull, D. and Traunmüller, H., "Perception of quantity in Estonian", *Proceedings, Fonetik 2000*, 85-88, 2000.
- [8] Eek, A. and Meister, E., "Foneetilisi katseid ja arutlusi kvantiteedi alalt (I). Häälikukestusi muutvad kontekstid ja välde", *Keel ja Kirjandus*, 11-12, 815-837, 904-918, 2003.
- [9] Meister, L., "Eesti vokaali- ja kestuskategooriad vene emakeelega keelejuhtide tajus ja häälduses", Ph.D. Dissertation, Tartu: Tartu Ülikooli Kirjastus, 2011.
- [10] Bondarko, L.V., "Fonetika sovremennogo russkogo jazyka", St.-Petersburg: University of St.-Petersburg, 1998.
- [11] Bohn, O.-S., "Cross-language speech perception in adults: First language transfer doesn't tell it all", in W. Strange [Ed], *Speech perception and linguistic experience. Issues in cross-language research*, Baltimore: York Press, 279-304, 1995.
- [12] McAllister, R., Flège, J. E. and Piske, T., "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian", *Journal of Phonetics*, 30, 229-258, 2002.
- [13] Meister, L. and Meister, E., "Perception of the short vs. long phonological category in Estonian by native and non-native listeners", *Journal of Phonetics*, 39, 212-224, 2011.
- [14] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer", (Version 5.2.25) [Computer program] Online: <http://www.praat.org>, accessed on May 15, 2011.