DURATIONAL PATTERNS OF THE TRISYLLABIC FOOT IN SOIKKOLA INGRIAN

San Francisco State University

A thesis submitted to the faculty of

San Francisco State University

In partial fulfillment of

The Requirements for

The Degree

Master of Arts In Linguistics

by

Irina Brodskaya

San Francisco, California

December 2015

Abstract

This paper presents a first ever quantitative analysis of the trisyllabic foot durational patterns in the Soikkola Ingrian dialect, an understudied Finnic variety of North-Western Russia on the verge of extinction. This dialect has a phonemic two-way durational contrast in vowels and a typologically rare three-way durational contrast in consonants. The acoustic analysis comprises the durational measurements of all foot nucleus segments, in a dataset of 3300 tokens collected from five speakers. The aim is to disentangle the effects of (1) the phonemic length contrasts, (2) the isochronic prosodic tendency to maintain the cumulative foot duration, and (3) the recently developed phonetic reduction in the non-initial syllables. The results show the preservance of the ternary contrast in consonants and the binary contrast in the first syllable vowels. However, the isochronic and the reductive tendencies have affected the durational contrast in the non-first vowels. The analysis shows that Soikkola Ingrian is undergoing a shift from the “northern” type of Finnic languages, where the second syllable vowels are contrastive in duration regardless the structure of the first syllable, to the “southern” type, where the duration of the second syllable is not phonemic and highly depends on the structure of the first one.

Acknowledgments

I would like to express my sincere gratitude to my graduate advisors Prof. Troi Carleton and Prof. Jenny Lederer for introducing me to Linguistics, for their encouragement, continued guidance, recommendations, and constructive suggestions during the planning and development of this research work.

I also want to thank Clara Cohen for her help in learning statistical analysis techniques.

I would like to extend my gratitude to Natalia Kuznetsova, who gave me the permission to use her data, and whose insightful comments helped me improve this manuscript. I am also grateful to Elena Markus for sharing her data with me.

I do not have enough thanks for my friends Piotr Achinger and Andrey Indukaev for their technical support on this project.

Finally, I wish to thank my husband Alexander Shapiro and my parents for their eternal support and encouragement throughout my study.

Contents

Abstract 2

Acknowledgments 3

1. Introduction 5

2. Geographical Position and Phonology of Soikkola Ingrian Dialect 7

3. Background 12

5. Methodology 18

6. Results 23

6.1 Phonemic Durations 23

    Experiment 1. Duration of C2 23

    Experiment 2. Binary contrast in the vowels of first syllable (V1) 26

    Experiment 3. Binary contrast in the vowels of second syllable (V2) 28

    Experiment 4. Interaction of V1 and V2 in structures with secondary geminates 32

    Summary 35

6.2. Phonetic realization: Segment durational compensation 36

    Experiment 5: Influence of C2 length on Long V1 Duration 36

    Experiment 6: The duration of vowels in feet with singletons 38

    Summary 40

7. Interpretation 42

8. Conclusions 45

9. Bibliography 47

APPENDIX I. 51

    A. Map of Europe 51

    B. Soikkola Peninsula 51

APPENDIX II. Words Analyzed 52

    Wordlist A: Trisyllables with Primary and Secondary Geminates 52

    Wordlist B: Trisyllables with Singletons 54

APPENDIX III: Linear Regression Models 55

1. Introduction

The present paper seeks to contribute to the study of Finnic prosody by presenting results of a quantitative research on duration patterns in the nearly extinct contemporary Soikkola Ingrian dialect. The aim of this investigation is to establish how foot prosodic structure (i.e., the alternation of short and long elements in a word) in a trisyllabic foot determines the realization of durational contrasts in vowels and consonants.

Soikkola Ingrian has a typologically rare three-way durational contrast in consonants: primary geminates (long duration) vs. secondary geminates (half-long duration) vs. singletons (short duration); in vowels it has a binary durational contrast (short vs. long vowels).

These two contrasts interact with two general phonetic phenomena: *foot isochrony* and *reduction*. The tendency towards the foot isochrony in the context of Finnic languages, is understood as a unification of the cumulative foot duration, regardless the number and duration of individual segments in a foot (Kuznetsova 2009a, 2009b, Markus 2010). Some major effects of the Finnic foot isochrony are known from other studies (Lehiste 1960, 2008), but not always confirmed for individual Finnic varieties (Lehtonen 1970). In the Ingrian prosody, this tendency creates the following constraints (Kuznetsova 2009b, 2012; Markus 2010, 2011):

1. In structures with a first short syllable, the second vowel has undergone a prosodic lengthening (* kat**a** -> kat**a**:[1]).
2. A long vowel shortens the following geminate.
3. A long geminate shortens the preceding long vowel.
4. Polysegmental and polysyllabic shortening (the more elements in a word, the less the duration of each of them)[2].

The reductive tendency has been running for the last few decades; it is manifest in the reduction of non-first vowels in long structures (Kuznetsova 2009a, 2009b, Markus 2010, 2011). Reduction results in the loss of some quantitative contrasts in the non-first syllables.

The phonetics of Ingrian is under-investigated. The prosodic system was previously researched in an experimental study only regarding a disyllabic foot with an open second syllable

1 Key: C:ˑ – full (primary) geminate, C: – half-long (secondary) geminate, C – singleton; V – short vowel, V: – long vowel, V1V2 - diphthong, R – sonorant. NOM – Nominative Case; PART - Partitive Case; ALL – Allative Case; PRS – Preset Tense; PST – Past Tense; SUP – Suppine; 1 – 1st person; 2 – 2nd person; 3 - 3rd Person; SG – Singular; PL – Plural, CNG – Connegative.

2 This research does not allow the investigation of this feature, since the dataset contains only the trisyllabic words.

(Kuznetsova 2009b; Markus 2010, 2011). However, there is evidence that durational patterns of di-and trisyllabic feet differ.

The quantitative contrasts in non-first syllables have some constraints. In disyllabic feet, a secondary geminate can be followed only by a long V2. This is conditioned by the historical occurrence of secondary geminates: it is generally considered (Porkka 1885; Laanest 1966, 1986) that originally they developed out of singletons *before long vowels and diphthongs* (*kan**a**: 'hen:SG:PRT' > [ka**n**:**a**:], *min**ua** 'I:SG:PRT' > [mi**n**:**ua**]). However, in some works (Porkka 1885, Kettunen 1930, Sovijärvi 1944) it was mentioned that in trisyllables secondary geminates could occur before short vowels (vu:t**a**va, "leaking:NOM"), as well as before long vowels (su:t:**i**:ma "judge:1PL"). If this combination truly exists, this would contradict the generalization of the secondary gemination conditions.

Thus, the object of this study is limited to the trisyllabic foot with an open third syllable. In particular, I answer the following questions:

1. Is the ternary opposition of consonants manifest between first and second syllables?
2. Is the binary contrast in vowels manifest in the first syllable?
3. Is the binary contrast in vowels manifest in the second syllable after both primary and secondary geminates?
4. What are the correlations between the duration of different segments in the foot on the phonetic level?

Terminology. In this subsection I explain terminology and clarify the terms that I use throughout this paper.

Foot. The concept of foot is crucial in order to describe quantitative contrasts in Ingrian. In this work I follow the definition used by Kuznetsova (2009b, 2012). Foot is defined as a phonological unit larger than a syllable and smaller than (or equal to) a phonetic word; interrelations between segments (consonants and vowels) are tighter within a foot than between segments across the foot border. In addition, each foot has one rhythmic accent that always falls on the first syllable. The most typical foot type in Ingrian is disyllabic, but mono- and trisyllabic feet are also present.

Foot nucleus is a convenient term for description of prosodic mechanisms within a foot. It refers to a sequence of segments starting with the vowel of the first syllable (V1), and ending with vowel of the second syllable (V2). The consonant between them is referred as C2. For example, the foot with a structure C**VCV**CV is referred as **VCV**. The term *foot nucleus* is useful for describing so-called foot languages, since it refers to crucial elements of the durational opposition and interaction. Although there are languages where initial syllable consonant does play a role in quantitative contrasts (Topintzi, in press), the majority of scholars who study Finno-Ugric languages consider this not to be the case for this group (Prillop, 2013). The concept of mora helps us understand this idea.

Mora is a term that refers to a syllable weight (light vs. heavy vs. super-heavy). Typically, the syllable onset has no mora; an open syllable with a short vowel has one mora; a syllable with a long vowel, and a syllable with a coda have two moras; trimoraic syllables (with a long vowel and a coda) are also possible in some languages. How many moras form a syllable of a certain weight depends on a specific language (Oostendorp, 2005). Traditionally, in Soikkola Ingrian, there is a distinction between two initial syllables types (Kuznetsova, 2012, p.55):

1. light
   - open syllables with a short vowel, for example VCV: ([kata:] "roof:IMP:2SG")
2. heavy
   - closed syllables with short vowels: for example VRCV ([palkale] "wage:ALL", VC:ˑV, (kat:ˑila "boiler:NOM")
   - open syllables with long vowels or diphthongs: for example V:CV: ([sa:tama] "accompain:1PL")

2. Geographical Position and Phonology of Soikkola Ingrian Dialect

The Ingrian language belongs to the Finnic branch of the Finno-Ugric group, Uralic family. It is spoken in Ingria, (see the map in Appendix I), a region of Russia on the Eastern shore of the Baltic sea, the Leningrad region (Kuznetsova 2009a). There are four dialects of the Ingrian language: Lower-Luga, Soikkola, Hevaha, and Oredezh; the latter two are now extinct. This paper focuses on the Soikkola dialect spoken in the Western part of Ingria (Soikkola peninsula, see the map in Appendix I). According to Kuznetsova (ibid.), the Soikkola dialect is subdivided into two main branches: *Northern* and *Southern* with an intermediate *Central* variation spoken in the Vi:stina (Vi) village, situated between two dialect areals. Consider Fig. 1 (The map is adapted from Acta Linguistica Petropolitana, 2012; the key is adapted from Kuznetsova 2009a).

Sociolinguistics situation. The Soikkola dialect is nearly extinct. In 2006 there were 67 fluent speakers attested in the area; there is no data about the current number of speakers, but now the language is even closer to extinct and numbers just a few dozens of speakers. Ingrian is not a language of everyday communication. Speakers of Ingrian are scattered among different villages in the area. After the World War II the Ingrian language underwent a drastic change in its social status. The post-war language policy enforced Russian language upon the speakers of all other national minorities, which resulted in a rapid decrease of the communication in Ingrian (Kuznetsova, at al., 2015).

Taxonomy. Ingrian is a part of the Finnic branch of the Finno-Ugric group, that also includes such languages as Veps, Karelian, Votic, Estonian, Finnish, and Livonian. They developed from the Proto-Finnic language, where, aside from the quality, the phonemic differences between sounds were manifested in the duration; after the proto-language split, the different Finnic languages developed unique prosodic systems. They form a continuum with respect to the complexity of the stem alternations. On the one extreme of this continuum there are languages with the simplest prosodic systems, such as Veps that have almost lost the sound duration oppositions (Viitso 1961, Kuznetsova 2015a). On the other extreme, there are languages with the most convoluted prosodic systems and quantity oppositions, such as Livonian and Estonian (Kuzntesova, 2015).

Phonemic inventory. There are 8 vowels in Ingrian; they are represented in Table 1 (Kuznetsova, 2009a, p. 169). Phonemes that are not common for all varieties are put in parentheses.

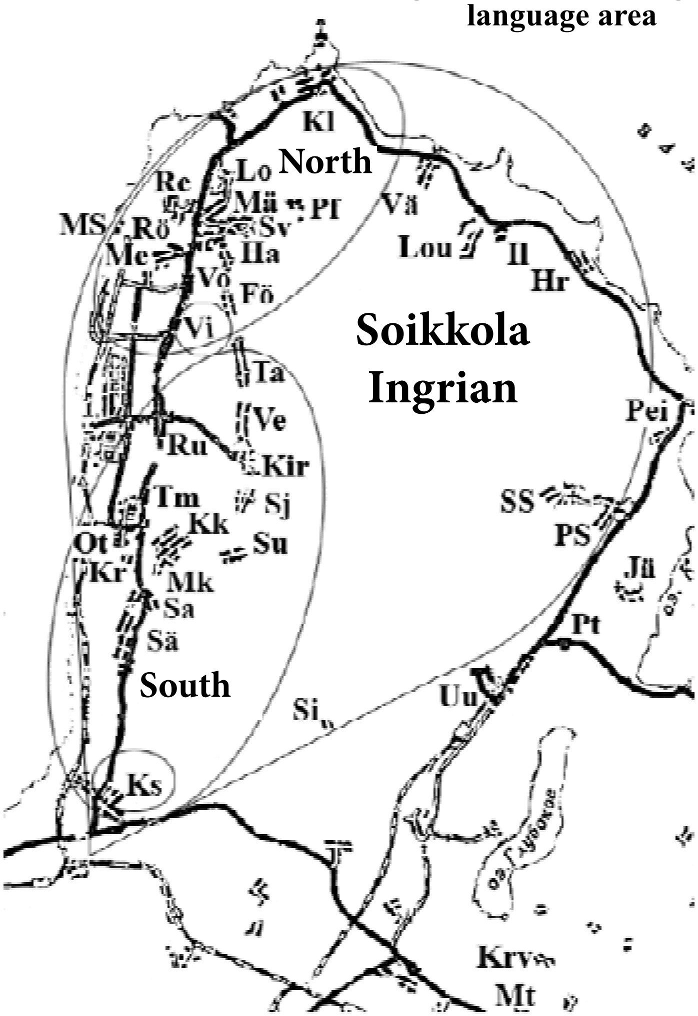
Table 1: Vowels in Soikkola Ingrian

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	i iː	ü (üː)		u uː
Middle	e (eː)	ö (öː)		o (oː)
Low	ä (äː)		a aː	

The Soikkola Ingrian phonemic inventory also includes about 20 diphthongs.

Ingrian has front-back vowel harmony with neutral *i* and *e*, which can participate in the

Figure 1: Soikkola Ingrian language area



Key:

Ha – Hamːˑala
Hr – Harkːˑola
Kk – Koskina
Kr – Kr´aːsnagorkːa
Kri – Krivul'ka
Krv – Korvatːina
Ks – Koskisenkülä
Lo – Loka, Lokove
Mä – Mätːˑüisi
Me – Metsäkülä
Ot – Otsave
Re – Repːˑola
Rö – Rölːˑölä
Ru – Ruːtsia
Sa – Saːrve
Sä – Säːtinä
Sj – Sujːa
Su – Sutela ~ S´utel(ə)
Ta – Tarinːˑaisːˑi
Tm – Tamːˑikontːˑu
Vä – Väːräoja
Vi – Viːstina

forms with both front and back vowels.

Cardinal consonants are represented in Table 2[3]. Cardinal phonemes are boldfaced, peripheral phonemes (found in borrowed words and/or expressive lexicon) are not marked; *(ʃ ʃː)* are attested only in some Southern idiolects (Kuznetsova, 2009a, p. 170).

3 This is the most up-to-date phonemic inventory; however, it describes the language on the accent, not on the segmental level, therefore reflects only cardinal phonemes (singletons and primary geminates).

Table 2: Consonants in Soikkola Ingrian

	Labial	Alveolar	Palatalized	Velar and Laryngeal
Plosive Oral	**p p**:	**t t**:	*t' t'*:	**k k**:
Nasal	**m m**:	**n n**: [n/ŋ]	*n' n'*: *[n'/ŋ']*	
Fricative	**f f**:	**s s**:	*s s*:	**h h**:
		(ʃ ʃ:)		
Lateral Approximant		**l l**:	*l' l'*:	
Trill		**r r**:	*r' r'*:	
Affricate			*ɕ ɕ*:	
Glide	**v v**:		**j, j**:	

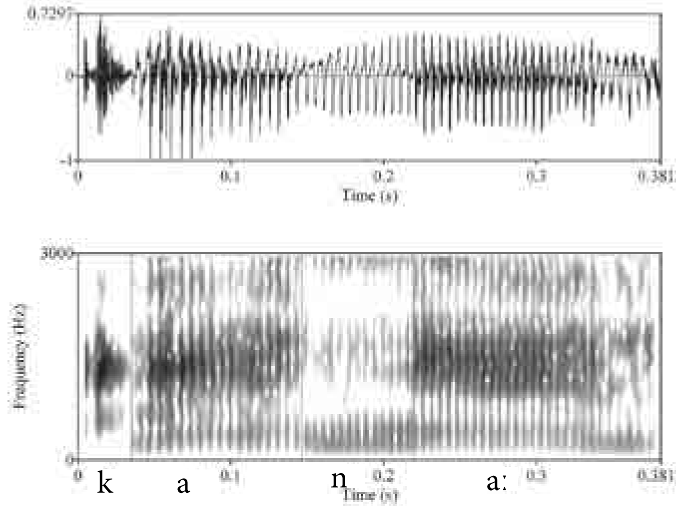
Ingrian has a complex system of duration patterns; on the phonological level there is the binary contrast in vowels and the ternary contrast in consonants. Phonemes /p/, /t/, and /k/ are contrastive not only in length (long vs. half-long vs. short), but also in voiceness (voiced vs. voiceless) in some positions. In those cases singletons are voiced, and consonants of other durations are voiceless ([ta**b**aː] "kill:IMP:2SG" vs. [ta**p**ːa] "catch:PRS:3SG"). This reinforces the durational opposition, serving as a secondary cue. In this work I chose to use the phonological transcription for singletons to stay consistent with the purpose to show the durational oppositions. For example, [ta**b**aː] is transcribed as [ta**p**aː]. Except for the case with the stops, all other phonemes are contrastive in length.

The three-way durational opposition is illustrated in the following example:

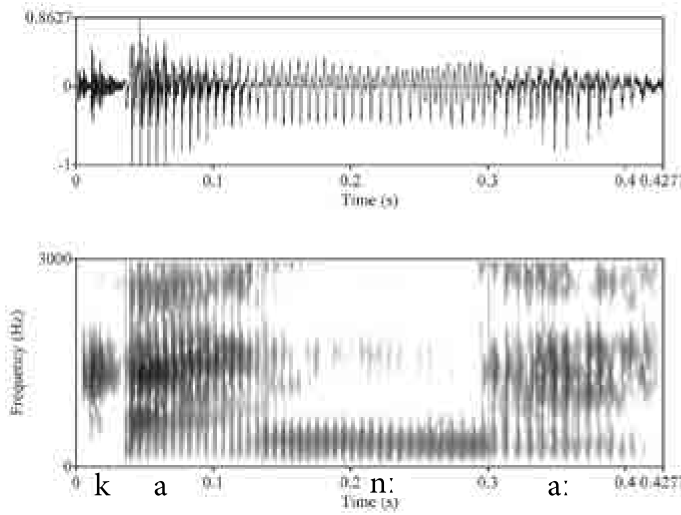
[kana] "hen:NOM" vs. [ka**n**ːaː] "hen:PART/ILL vs. [li**n**ːˑaː] 'town: "PART/ILL"

Figures 2—4 provide examples of spectrograms for each word, pronounced by AJF, a female born in 1933 in a Hammaala (Ha) village (see fig.1) (Kuznetsova 2015a).

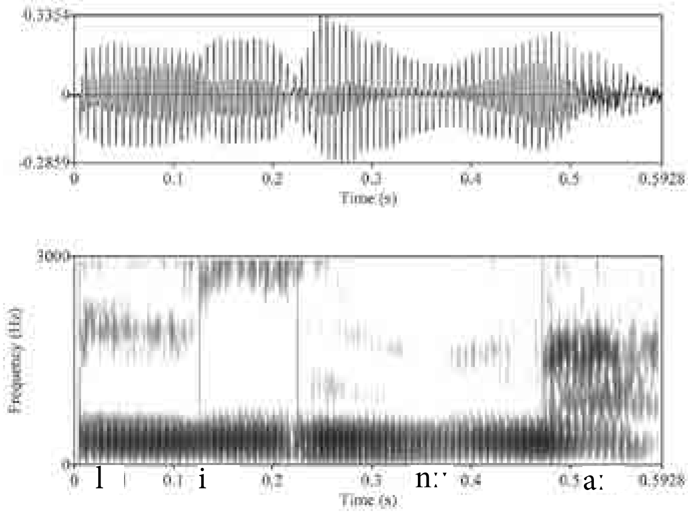
Figures 2—4: Spectrograms of the Tokens Pronounced by AJF



[kanaː]



[kanːaː]



[lin:ˑa:]

3. Background

This section gives an overview of previous studies dealing with prosodic features of Ingrian language, and specifically quantitative research. Ingrian prosody has been addressed in the following research papers. A ternary contrast in consonants was first observed by Porkka (1885) and later confirmed in Sovijarvi (1944), who conducted the first instrumental study on Ingrian prosody; another experimental study was conducted by Laanest (1978).

In more recent research, Kuznetsova (2009b) provides some experimental measurements on disyllabic foot, and confirms the ternary contrast as well. Gordon (2009) analyzed a spontaneous speech recording made in 1964 in the Southern part of the Soikkola peninsula. The analysis he provides to illustrate quantity oppositions in Ingrian is not consistent with the claim that the durations of short and long geminates differ. Markus (2010) conducted research on her own data from five speakers collected in two villages located in the Northern and one in the Central part of the Soikkola peninsula. Her research is based on isolated sentences that contain target words sentence- and phrase finally. The author provides solid evidence that in disyllables ternary contrast in consonants is manifest in duration. However, in trisyllables this contrast was not confirmed.

The methods of data collection and analysis and the difference in the origin of data could have triggered the discrepancy in the results reported by Gordon (2009) and Markus (2010). The durations of corresponding segments in the spontaneous speech and in isolated sentences might differ (for example, Arvaniti 2012). Furthermore, in Gordon (2009) the author tested the structures of a different number of syllables together, while in Markus (2010) the researcher conducted her analysis on disyllabic and trisyllabic words separately. Polysegmental shortening as an isochronic effect could have shortened the duration of geminates in long structures, which might have resulted in the lack of difference between the corresponding segments in structures of different length. Finally, the Southern Soikkola dialect is spoken in the area, which is geographically closer to the Lower-Luga Ingrian dialect. The opposition of full and short geminates there was completely lost, and it might have influenced that particular idiolect.

Markus (2011) conducted another study, specifically on disyllabic foot with an open second syllable and an intervocalic consonant. She concludes that on the phonetic level, there are three types of vowel duration, but only two of those are phonemic. Phonologically short and long vowels can appear in either the first or second syllable. The third type of vowels in older research called *half-long* are found to be longer than long vowels, and referred to as *overlong*. Their distribution is complimentary: those vowels occur only in structures with a light first syllable (VCV foot nucleus), which means that it is rather a phonetic representation of a short vowel, than a phonologically different entity. Thus, in disyllables the following durational patterns were attested:

1. Two-way durational contrast in vowels in the first syllable: [s**a**vi], “clay:NOM” vs. [s**aː**vi], “tub:NOM”.
2. Three-way durational contrast in consonants between first and second syllables: [ta**p**aː] “kill:IMP:2SG” vs. [ta**p**ːa] “catch:PRS:3SG” vs. [ta**p**ːˑa] “kill:INF”
3. Two-way durational contrast in vowels after primary geminates: [saːtːˑ**a**ː] “accompany:PRS:3SG” vs. [vuːtːˑ**a**] “year:PART”.
4. Vowels after singletons undergo a prosodic lengthening in words with a short first syllable: [kat**a**ː], “roof:IMP:2SG”.
5. Secondary geminates are always followed by long vowels (no durational contrast): [makː**a**ː] “sleep:IMP:2SG”, [tukː**a**ː] bring:IMP:2PL

The author notes that this contrast in duration of short and long V2s is quite small: 20 to 30 ms; however, in low vowels ([a] and [ä]), there is an audible difference in quality: short vowels are reduced and sound like a *schwa*, and the long ones sound like full vowels. The binary contrast in non-first vowels is shifting from *long* vs. *short* vowels to *full* vs. *reduced* vowels, which indicates the transition between the “northern” type of prosodic system, which is closer to Finnish, to the “southern” type, which is closer to Estonian.

With respect to consonants, Markus confirmed that disyllabic foot has durational contrast the intervocalic position: singletons vs. primary (long) geminates vs. secondary (short) geminates. On the phonetic level they are realized as five distinctive categories, following the isochronic tendency (from the shortest to the longest phonetic realization): singletons < short geminates preceded by long vowels < short geminates preceded by short vowels < long geminates preceded by a long vowels < long geminates preceded by short vowels.

Short vowels after secondary geminates. In a number of papers Kuznetsova (2009a, 2009b, 2012) described the system of foot quantity, using foot accents. The accent system accounts for quantity on the suprasegmental level, relating to the stress and quantitative balance of segments in a foot, which helps more accurately describe prosody of segments altogether, not just individual durations on the segmental level.

Initially, the author accounted for three types of segments quantity combinations in a foot nucleus, given that the first syllable can be either short or long (Kuznetsova 2012, p.48), namely:

1. Primary geminate + long vowel (VCːˑVː) (makːaː, matːaːla)
2. Primary geminate + short vowel (VCːˑV) (saːtːˑaː, muːtːˑiːma)
3. Secondary geminate + long vowel (VCːVː) (saːtːˑaː, suːtːiːma)

However, in one of her more recent papers (Kuznetsova 2015b), the author points at the existence of one more combination, that was not investigated before:

4. Secondary geminate + short vowel (VCːV) (vuːt**a**va)

This structure was mentioned by Laanest (1966), Porkka (1850) and Sovijarvi (1944), however they have not discussed this phenomenon in detail. It has been generally considered that in Ingrian, primary geminates were inherited from the Proto-Finnic language, while secondary geminates later developed from singletons before long vowels and diphthongs (*kan**a**ː ‘hen:SG:PRT’ > [ka**n**ː**a**ː], *min**ua** ‘I:SG:PRT’ > [mi**n**ː**ua**]). For this reason Markus (2011) supposed that a short vowel after a primary geminate would not be possible (ibid., p.105), and indeed, in the disyllabic structures of this type it was not attested. Therefore, the question of whether the forth combination in the foot nucleus is present in the Soikkola Ingrian dialect, is crucial.

Table 3 summarizes the patterns of the binary contrast in vowels, common for di- and trisyllabic words; the uncommon word structure (4) is shaded.

Table 3: Binary Contrast in Second Vowel after Three Kinds of Consonants

V2	# of syllables	Consonant Length		
		Singleton	Primary Geminate	Secondary Geminate
Long	2	[kataː] roof:IMP:2SG	[saːtːˑaː] accompany:PRS:3SG	[makˑaː] sleep:IMP:2SG
	3	[otaːma] take:1PL	[muːtːˑiːma] change:1PL	[matˑaːla] low:NOM
Short	2	[saːta] accompany:IMP:2SG	[vuːtːˑa] year:PART	[suːtˑiːma] judge:1PL
	3	[saːtama] accompan:1PL	[katːˑila] boiler:NOM	[vuːt**a**va] leaking:NOM

Foot isochrony in Finnic languages. The tendency towards foot isochrony is claimed to be a typical feature of some Finnic languages, including Ingrian (Lehiste 2008 regarding Livonian, Lehiste 1960 regarding Estonian, Kuznetsova 2009b, 2012, and Markus 2010, regarding Ingrian). This is a controversial term that means a few different things. First, it was used within a theory that suggested a regularity in time alignment of speech elements (Pike, 1945; Abercrombie, 1967). Languages were divided into two types: stress-timed and syllable-timed. Stress-timed means that stress should occur in equal intervals, irrespective to the number of segments in between. Syllable-timed means that syllables occur in equal time intervals, irrespective to stress.

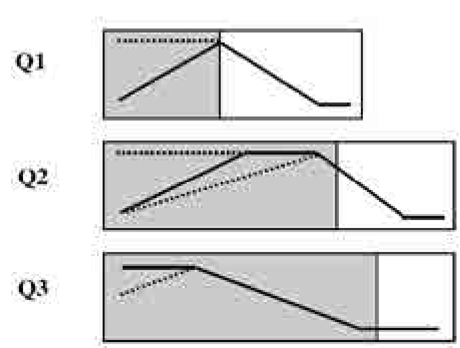
Later on, Wiik (1991) suggested to distinguish the third type of languages called *foot-timed*. In the systems with this type of isochrony, the stress should occur in equal intervals, just like in stress-timed languages, but the domain that conditions the occurrence of this effect is a foot. The total duration of a foot tends towards the unification irrespective to the number of elements; segments alter each other's durations within a foot. For example, in Estonian this is manifest as follows. There are three contrastive syllables' lengths (referred as Q1, Q2, and Q3, from the shortest to the longest duration) in both vowels and consonants. However, those durations are phonemic only in the first syllable. The duration of the second syllable varies according to the length of the first syllable: the shorter the first syllable, the longer the second syllable; the longer the first syllable, the shorter the second syllable (Wiik, 1991, Lehiste, 1960; Krull, 1999; Prilopp, 2013). Table 4 summarizes the phonemic contrasts and their phonetic representation in the Estonian language:

Table 4: Phonemic Contrasts in Estonian and their Phonetic Representation

Quantity	#	Phonemic foot nucleus structure	Phonetic representation
Q1	1	VCV	VCV:ˑ
Q2	2	V:CV	V:CV:
	3	VC:V	VC:V:
	4	V:C:V	V:C:V:
Q3	5	V:ˑCV	V:ˑCV
	6	VC:ˑV	VC:ˑV
	7	V:ˑC:ˑV	V:ˑC:ˑV

The table shows that on the phonemic level, the shortest first syllable in Q1 is followed by the longest V2; in Q2 the half-long first syllable is followed by a half-long V2; the longest first syllable is followed by the shortest V2. This is illustrated in Figure 2, where each bar represents the first (gray) and the second (white) syllables durations. Thus, the ternary contrast in Estonian does not stand by itself: the duration is accompanied by secondary cues that manifest in the duration of subsequent syllables. Other secondary cues are the intensity and the pitch contour. Figure 5 summarizes schematically the duration ratios and the pitch contour movement in the three quantities (Asu, 2004).

Figure 5: Ternary Durational Contrast in Estonian



In regard to Ingrian, foot isochrony was mentioned by Sovijarvi (1944, p 11, 15-16), described by Kuznetsova (2009b p.37, 2012 p. 50, 2015a p. 200, 2015b), and by Markus (2010, p. 48, 50). It manifests itself in the tendency toward a unification of the cumulative foot duration, and durational compensation. In the VCV structure, this effect is the most prominent: the second vowel undergoes a substantial lengthening, for example: /kat**a**/ > [kat**a**ː] “roof:IMP:2SG”. The distribution of this vowel type is complimentary, at least in disyllabic foot: long V2s occur only in structures with a light first syllable (VCV foot nucleus), therefore it is not considered as a separate phoneme. It could be an allophone of either short or long vowels; since this paper describes the durational distribution on the phonetic level, I find it most convenient to refer to them as to *long vowels*.

Another isochronic effect that was attested in disyllables, is the tendency towards the compensatory change of segments’ duration. It manifests itself in the difference in average durations of consonants, that depends on which vowels they follow. A primary geminate is longer after a short vowel, and it is shorter after a longer vowel. The same effect applies to secondary geminates. For example, the difference in VCːV and VːCːV lies not only in the phonemic length of the V1, but also in the phonetic realization of the C2 duration. Thus, the three-folded phonemic contrast is phonetically realized in five different ways: singleton (VCVː, [kataː] < secondary geminates after long vowels (VːCːV, [tuːkːaː]) < secondary geminates after short vowels (VCːVː, [makːaː]) < primary geminates after long vowel (VːCːˑV, [vuːtːˑa]) < primary geminates after short vowels (VCːˑV, [kukːˑa]) (Markus, 2011).

A similar effect was found in the genetically close Finnish language, which, unlike Estonian, demonstrates the binary contrast in vowels and consonants in any position. Comparison of a series of minimal pairs that have durational differences in one phoneme, showed that while those words have differences in certain corresponding elements on the *phonemic level*, they also show the differences on the *phonetic level*, in the elements other than those with the phonemic contrast. For example, the comparison of CV**C**VCV and CV**C**ːVCV structures shows that not only do intervocalic (boldfaced) consonants differ in quantity, but also does the following vowel. V2 in CVCVCV is 25 ms longer than V2 in CVCːVCV. The difference on the phonetic, but not phonemic level, can be treated as a secondary cue that reinforces the durational contrast (Aoyama, 2001).

Thus, Finnish and Estonian, having the same origin, demonstrate differences in prosodic systems. In Estonian, the duration of the second syllable vowel differs, but it is not phonemic; it is rather complimenting the duration of the first syllable. This is a "southern" type of language (also called *accent* language). Finnish, is related to the "northern" type (also called *quantity* language): the segments do not depend on one another, long and short vowels and consonants can appear in any position within a foot. Analysis of the disyllabic foot has shown that the contemporary Soikkola Ingrian is in transition between southern and northern types of languages: the durational contrast in non-first syllable vowels still remains, like in Finnish; however, it is gradually getting neutralized, like in Estonian. The durational contrast is shifting from both syllables towards the first syllable only.

Thus, the current analysis aims to discern which of those isochronic effects govern the trisyllabic foot structure, and how far the process of transition between northern and southern types of languages has gone.

5. Methodology

In this section I show how the questionnaires for this study were designed, and the methods of data collection.

Data was collected by Natalia Kuznetsova, Elena Markus, and the author in 5 villages in the Kingisepp district, Russia during three field trips in summer 2014 and spring 2015. All the speakers are females who were born in 1920s and the 1930s. Speakers 1 through 4 belong to the Northern sub-variety area. Speaker 1 and 2 (from Mä:t:üisi)[4] have learned Ingrian in early childhood as a first language, speaker 3 (from Ha:m:ala) learned Ingrian at the age of 5 years old, speaker 4 (from Rep:ola) has learned Ingrian as a first language, however she had a long break from using it during her lifetime. She also might have some irregularities since she learned Ingrian from her parents who lived in a different village than she lives now, therefore she might have developed some accent that is hard to track. For all these reasons I expect differences in idiolects of speaker 1 and 2 in comparison to speakers 3 and 4. Speaker 5 stands out, since she lives in the Viistina village, which is situated on the border between Southern and Northern sub-varieties.

Questionnaire 1. The word list contains word structures with three syllables and a geminate (primary and secondary) on the border of first and second syllables. The consonants of the second syllables are limited to stops (/p/, /t/, /k/). Words may contain a short V1 or a long V1, as well as short V2 or a long V2. Besides gemination and vowel length, we control such variables as V1 as a single sound vs. a diphthong, presence/absence of a sonorant in the coda of the first

4 See the map (Fig. 1)

syllable. This way the list of target words contains 17 possible categories, each of which has the listed properties in all possible combinations.

Table 5 provides examples for each word type. The problematic structures (secondary geminate+short vowel) are shaded. It is noteworthy that the examples 5 and 12 make a minimal pair, which can speak in favor of the ternary contrast in consonants. For the full list of target words see Appendix II.

Table 5: Questionnaire 1. Primary and Secondary Geminates in Trisyllables

#		Diphtong/ sonorant	Foot Nucleus	Example	Gloss	N tokens
1	Primary Geminates		VCːˑV	katːˑila	boiler:NOM	105
2			VCːˑV:	katːˑiːma	cover:1PRS.PL	109
3			V:CːˑV	sa:tːˑaja	guide:NOM	120
4			V:CːˑV:	va:tːˑiːma	get.dressed:1PRS.PL	108
5		+ Diphthong	V1V2CːˑV	voitːˑeli	fight:3PST	113
6		+ Diphthong	V1V2CːˑV:	toitːˑiːma	eat:1PRS.PL	120
7		+ Sonorant	VRCːˑV	tarkːˑoja	smart:PART.PL	127
8		+ Sonorant	VRCːˑVː	harkːˑaːma	step:1PRS.PL	113
9	Secondary Geminates		VCːVː	matːaːla	low:NOM	180
10			**VːCːV**	**naːtːala**	**stove.bottom:NOM**	**111**
11			VːCːVː	laːtːiːma	get.ready:1PRS.PL	101
12		+ Diphthong	**V1V2CːV**	**voitːeli**	**smear:PAST.3SG**	**173**
13		+ Diphthong	V1V2CːVː	hoitːiːma	beware:1PRS.PL	189
14		**+ Sonorant**	**VRCːV**	**märkːeni**	**rot:3SG.PST**	**168**
15		+ Sonorant	VRCːVː	kerkːiːmä	be.on.time:1PRS.PL	180
16		+ Sonorant	**VːRCːV**	**piːntːara**	**seedbed:NOM**	**183**
17		+ Sonorant	VːRCːVː	väːntːiːmä	turn.around: 1PRS.PL	203

The experiment was conducted in the following way: each test word was asked in a sentence final position. This way all the contrasts are articulated prominently, and we do not run a risk of having a vowel reduction triggered by the following word in a sentence or a phrase. Thus, an example of a carrier sentence is: “Saunaas ono suuri **kattila**” – “In the sauna there is a big **boiler**”, where *kattila* is the target word.

Each category contains five words of the same type. After two field trips we have got each speaker pronouncing a sentence at least seven times, so that in total we get each word type pronounced 35 times. From four speakers we collected at least 595 utterances of each. Unfortunately, the fifth speaker could only participate in a part of this experiment; therefore data collected from her are limited to 3—4 utterances per word, and, in total, from her we recorded only 360 utterances.

Questionnaire 2. The list of word structures used for this questionnaire repeats the system used in Questionnaire 1, however this time, there is a singleton on the border of first and second syllables. There are only five such foot types possible. Table 6 provides examples for each word type:

Table 6: Questionnaire 2. Singletons in Trisyllables

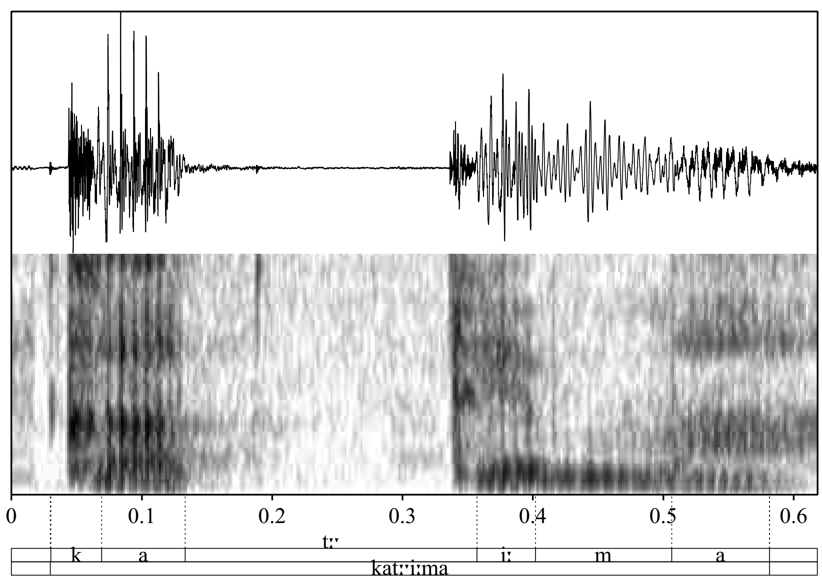
Diphthong/ sonorant	Nucleus structure	Example	Gloss	N tokens
	VCV:	ota:ma	take:1PRS.PL	67
	V:CV	sa:tama	accompany:1PRS.PL	43
+diphthong	V1V2CV	soitama	play:1PRS.PL	55
+sonorant	VRCV	palkale	wage:ALL	36
	V:RCV	pi:ntara:	seedbed:PART:PL	43

The questionnaire also contained five examples of each word type. For this subset, each sentence was pronounced only three-four times, therefore for each word type we collected at least 15-20 utterances. For this questionnaire, we managed to collect data from three speakers, (244 tokens). Therefore, in statistical analysis that aims to compare word structures with all types of consonants (singletons, primary and secondary geminates), smaller subsets with tokens from only three speakers are used, so that the subsets collected from different speakers are identical.

Technical procedures. All tokens were cut from the raw data files with the Adobe Audition 6.0 software. At this point, some tokens were omitted. These were to the cases when a consultant had accidentally mispronounced a word, or when some unwanted sounds had affected the record quality.

All tokens were concatenated and segmented with a forced aligner, and then hand-corrected, according to traditional phonetic criteria, in PRAAT acoustic analysis program (Boersma & Weenink, 2013). The segment borders were put based on the visual information drawn from spectrograms and oscillograms, and from the acoustic signal as well. For visual information I used gaps and subsequent bursts in the spectrograms for identifying stops, formant structure and sine wave change for identifying the borders of vowels and sonorants. In many tokens, the oscillogram that represents a word final vowel shows the friction produced by the airflow, not the vocal chords vibration. Acoustically this sounds as aspiration, and there is no audible sound of the vocal folds vibration. Therefore in many cases the final segment border was put before the sine wave ended. Figure 6 shows an example of a segmented token *kattiima.*

Figure 6: Token Sample Segmented in Praat



After the segmentation procedure, scripts, written in the computer programming language Python, were used in order to extract the durations. The output had the following variables speaker, word, segments' identities and durations (1st tier in the Praat segmented file), word duration (2nd tier in the Praat segmented file).

Analysis methods. A number of linear regression models were built in order to predict the relationships between the durations of the foot nucleus segments and their lengths. This type of analysis allows us to analyze imbalanced datasets, and also gives control over errors that come from the differences between words and idiolects, which is important with such a small number of speakers[5] (Johnson, 2011). Linear regression finds a linear relationship between a dependent numeric variable and a predictor variable.

Three sets of models were constructed in the R environment (R Core Team, 2015), using the *lme4* package (Bates et al., 2014). The first set of models analyzed V1 duration, the second set of model analyzed C2 duration, and the third set of models analyzed V2 duration. All models contained by-subject and by-word adjustments to the intercept. This means that the intercept (the point where the given line cuts the coordinate axis) is free to vary across the speakers and words (Baayen et al., 2008), reducing the error. The fixed effects were included through the backward elimination procedure. For this method a most general model that had all the predictors was built, and then each predictor one at a time, was dropped until the variables with statistically significant values remained. The AIC and the p-value were used as criteria to select the best model: the lower the AIC and the p-value, the better the model. A predictor is selected for the final model if both AIC and p value give a statistically significant result. If a given predictor improved the model fit, then it meant that there was statistical evidence that the independent variables had a more significant effect. This method of model building warranted that each predictor significantly improved the model fit. The test does not show exactly how the variable affects the target segment, but it shows whether or not it does. For all models statistically insignificant effects and interactions are removed, and subsequently I present only significant effects.

Subsequent linear regression analysis answers the question, to what extent the predictor affects the dependent variable, and whether this effect is significant. In this analysis I rely on the t value: once its absolute magnitude is equal or greater than 2, the difference between objects is considered statistically significant. However, it does not always mean that this difference is perceptible (Lehtonen, 1970). The minimal perceptibility threshold for speakers of English is 35-40 milliseconds (Lehiste, 1977). There is no data on the perception of Soikkola Ingrian speakers, but I assume that duration above 35-40 milliseconds must be perceptible for them.

Dependent variables are the elements of the foot nucleus. They are numeric values, and each data point represents one segment's duration in milliseconds (ms). Table 7 summarizes which variables we used for the analysis.

---

5 The author also ran the same analysis on each speakers' data separately, in order to see whether the tendencies found in speech of all five consultants would be attested in every individual speaker's idiolect; since the answer was positive, we can rely on results obtained from the analysis of five speakers together.

Table 7: Dependent Variables

Variable	Description
v1.dur	duration of the first syllable vowel (V1)
v2.dur	duration of the second syllable vowel (V2)
c2.dur	duration of the second syllable consonant (C2)

Predictors are displayed in Table 8. All of them are categorical, except for word duration (dur), which is numeric, and serves to account for an error that comes from the speech rate differences.

Table 8: Predictors

Variable	Categories	Description
v1length	short | long	length of V1
dip	yes | no	V1 is a diphthong, or a monophthong vowel
gem	prim | second | sing	type of the consonant in the second syllable
son	yes | no	structures with a sonorant ([l, r, n, m]) preceding the consonant in the second syllable, or without it
v2length	short | long	length of V2
dur	(numeric)	overall duration of a word, accounts for the speech rate

The exact models and subset descriptions for each test can be found in Appendix III.

6. Results

6.1 Phonemic Durations

Experiment 1. Duration of C2

In this experiment a multiple linear regression (mod 1, subset 1) was calculated to predict C2 duration based on **v1length, v2length, gem, son, dur, dip**. The statistical model estimation is summarized in Table 9. Since all values are significant, the final model contains all predictors. The interaction between V1 length and consonant length (**v1length:gem**) does not improve the model fit, which means that there is no significant correlation between the vowel length and the duration of a consonant that follows it. This prediction is not consistent with the finding that in disyllabic feet a long vowel shortens the following gemintes of both kinds (Markus 2011).

Table 9: Statistical model estimation: C2

		AIC	p value
C2 duration c2dur	dip	-11	<0.0001
	son	-61	<0.0001
	gem	-49	<0.0001
	v1length	-64	<0.0001
	dur	-398	<0.0001
	v2length	-5	<0.01
	v1length:gem	-3	0.6

In this experiment all possible structures with three lengths are tested. Table 10 shows the example tested, sorted according to the C2 length.

Table 10: Structures Tested: Primary Geminates vs. Secondary Geminates vs Singletons

	Foot Structure	Example	Gloss	N tokens
Primary Geminates	VCːˑV	katːˑila	boiler:NOM	915
	VCːˑVː	katːˑiːma	cover:1PRS.PL	
	V:CːˑV	sa:tːˑaja	guide:NOM	
	V:CːˑV:	va:tːˑiːma	get.dressed:1PRS.PL	
	VRCːˑV	tarkːˑoja	smart:PART.PL	
	VRCːˑVː	harkːˑaːma	step:1PRS.PL	
Secondary Geminates	VCːVː	matːaːla	low:NOM	1038
	VːCːV	naːtːala	stove.bottom:NOM	
	VːCːVː	laːtːiːma	get.ready:1PRS.PL	
	VRCːV	märkːeni	rot:3SG.PST	
	VRCːVː	kerkːiːmä	be.on.time:1PRS.PL	
	VːRCːV	piːntːara	seedbed:NOM	
	VːRCːVː	väːntːiːmä	turn.around: 1PRS.PL	
Singletons	VCVː	otama	take:1PRS.PL	244
	VːCV	saːtama	accompany:1PRS.PL	
	VRCV	palkale	wage:ALL	

The correlations of the variables are shown in Table 11. The *secondary geminate* level factor was chosen as a baseline to compare with *primary geminate* (**gemprim**) and *singleton* (**gemsing**). All correlations, except for **v2length** were statistically significant (t value > |2|). The results show that primary vs. secondary geminates are statistically significantly different in duration (t value = 16.524), as long as secondary geminates vs. singletons (t value = -16.494). Primary geminates depart from secondary geminates for 51 ms; singletons depart from secondary geminates for 78 ms.

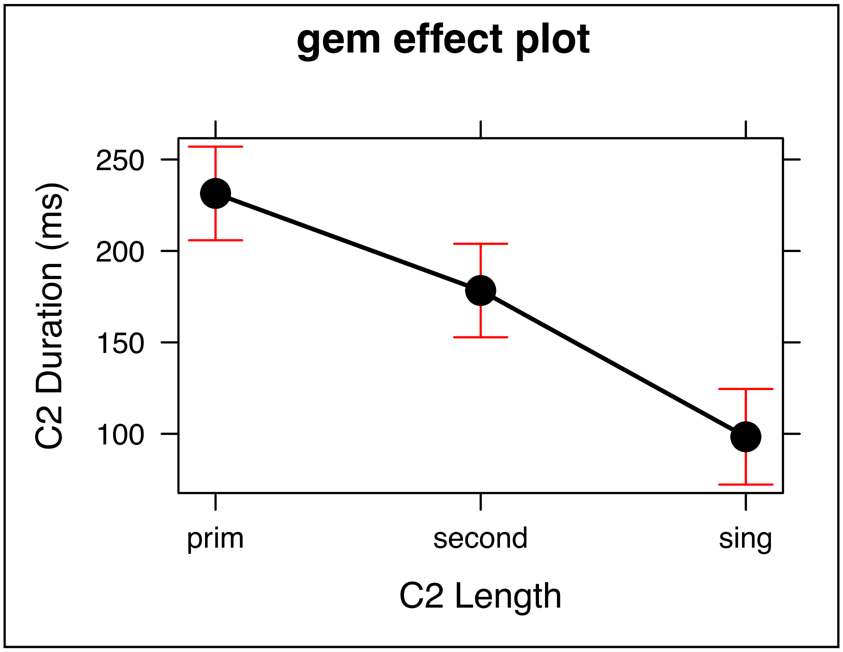
The model shows another significant correlation: V1 length. *V1 length:long* level factor was chosen as a baseline to compare with *V1 length:short* (**v1lengthshort).** According to this model, V1 length affects the C2 duration (t value = 8.268). I will return to this phenomenon in a subsequent analysis of a smaller data subset to test the isochronic effects.

Table 11: Correlation of Variables, C2 Duration

Fixed effects:	Estimate	Std.Error	t value
(Intercept)	67.465	16.028	4.209
sonyes	-32.092	3.255	-9.859
gemsing	**-78.804**	4.778	**-16.494**
gemprim	**51.666**	3.127	**16.524**
v1lengthshort	27.389	3.313	8.268
dur	0.140	0.007	20.913
v2lengthlong	12.109	8.762	1.382
v2lengthshort	18.106	8.396	2.157

Figure 7 plots the results of the correlation between C2 duration C2 length. Y-axis represents the duration in milliseconds (ms); x-axis represents C2 length factor levels: primary geminates (prim), secondary geminates (second), and singletons. Solid points represent the means; whiskers show the standard deviation from the mean.

Figure 7. Ternary Durational Opposition
in Consonants between First and Second Syllables



The durations of consonant with ternary contrast are, on average, 225 ms (primary geminates) vs. 175 ms (secondary geminates) and 100 ms (singletons). The difference between durations is above the perceptibility threshold of 35-40 ms, which, confirms that the three-way opposition is phonemic. This result differs from that reported by Gordon (2009), whose analysis shows the lack of secondary gemination, and Markus (2011), who did not confirm the ternary contrast specifically for trisyllables.

Experiment 2. Binary contrast in the vowels of first syllable (V1)

In this experiment a multiple linear regression (mod 2, subset 1) was calculated to predict V1 duration based on **v1length, v2length, gem, dur**. The statistical model estimation is summarized in Table 12. Not all values are significant, so in the final model I excluded the variables **dip** and **son.**

Table 12: Statistical model estimation

Dependent variables	Predictors Chosen	Model estimation	AIC	p value
V1 duration (v1dur)	v1length v2length gem dur	dip	+2	0.5
		son	+1	0.3
		gem	-21	<0.0001
		v1length	-142	<0.0001
		dur	-872	<0.0001
		v2length	-8	<0.001

Long vowels are chosen as a baseline and compared with short vowels. The correlations of the variables are shown in Table 13. All correlations were statistically significant (t value > |2|).

Table 13 presents the structures tested, sorted by V1 length.

Table 13: Structures Tested: Short V1 vs. Long V1

	Foot Structure	Example	Gloss	N Tokens
Short V1	VC:ˑV	kat:ˑila	boiler:NOM	897
	VC:ˑV:	kat:ˑi:ma	cover:1PRS.PL	
	VC:V:	mat:a:la	low:NOM	
	VRC:ˑV	tark:ˑoja	smart:PART.PL	
	VRC:ˑV:	hark:ˑa:ma	step:1PRS.PL	
	VRC:V	märk:eni	rot:3SG.PST	
	VRC:V:	kerk:i:mä	be.on.time:1PRS.PL	
	VRCV	palkale	“wage:ALL”	
	VCV	okama	take:1PRS.PL	
Long V1	V:C:ˑV	sa:t:ˑaja	guide:NOM	1300
	V:C:ˑV:	va:t:ˑi:ma	get.dressed:1PRS.PL	
	V:C:V	na:t:ala	stove.bottom:NOM	
	V:C:V:	la:t:i:ma	get.ready:1PRS.PL	
	V:RC:V	pi:nt:ara	seedbed:NOM	
	V:RC:V:	vä:nt:i:mä	turn.around: 1PRS.PL	
	V:CV	sa:tama	accompany:1PRS.PL	

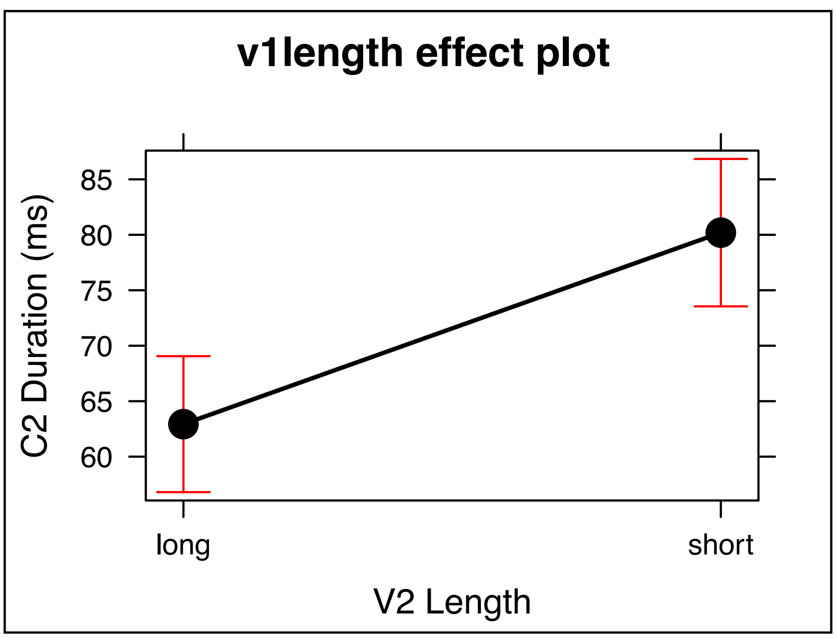
The correlations of the variables are shown in Table 14. The mean duration of the V1 in structures with *long V1* was compared with that of structures with *short V1* **(v1lengthshort).**

Table 14: Correlation of Variables, V1 Duration

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	47.160	9.072	5.198
v1lengthshort	**-64.642**	**3.232**	**-20.001**
gemsing	20.040	4.601	4.356
gemprim	-15.091	3.449	-4.375
dur	0.168	0.006	26.808

The results show that long vs. short vowels in the first syllable are statistically significantly different in duration (t value = -**20.001**). Short vowels depart from long vowels by 65 ms. Figure 8 plots the results of the correlation analysis between *V1 duration* and *V1 length*.

Figure 8. Binary Durational Opposition in Initial Syllable Vowels



In terms of perceptibility, 60 ms is significant, therefore, as predicted in previous research on diayllables (Markus 2011), V1 difference in duration is phonemic in trisyllables.

Experiment 3. Binary contrast in the vowels of second syllable (V2)

In this experiment a multiple linear regression with interaction (mod 3, subset 2) was calculated to predict V2 duration based on **v1length, v2length, gem, son, dur,** and the interaction **gem:v2length.** The subset contains only the tokens with primary and secondary geminates because structures with singletons have less variation and cannot be directly compared with other structures; feet with singletons will be tested separately in the subsequent analysis.

The statistical model estimation is summarized in Table 15. Not all values are significant, so in the final model I excluded the variable **dip.** Interestingly, the **v2length** variable effect is very small.

Table 15: Statistical model estimation

			AIC	p value
V2 duration v2dur	son v1length v2length gem dur	dip	+1	0.2
		son	-15	<0.0001
		gem	-20	<0.0001
		v1length	-35	<0.0001
		dur	-183	<0.0001
		v2length	-2	0.04

Table 16 shows the sample tokens analyzed, sorted by V2 length and C2 length.

Table 16: Structures Tested: Interaction of C2 Length and V2 Length

Geminate	V2 Length	Foot Structure	Example	Gloss	N Tokens
Secondary Geminate	Short V2	V:C:V	na:t:ala	stove.bottom:NOM	712
		V:RC:V	pi:nt:ara	seedbed:NOM	
		VRC:V	märk:eni	rot:3SG.PST	
	Long V2	V:C:V:	la:t:i:ma	get.ready:1PRS.PL	929
		VRC:V:	kerk:i:mä	be.on.time:1PRS.PL	
		V:RC:V:	vä:nt:i:mä	turn.around: 1PRS.PL	

Primary Geminate	Short V2	VC:ˑV	kat:ˑila	boiler:NOM	754
		V:C:ˑV	sa:t:ˑaja	guide:NOM	
		VRC:ˑV	tark:ˑoja	smart:PART.PL	
	Long V2	VRC:ˑV:	hark:ˑa:ma	step:1PRS.PL	720
		V:C:ˑV:	va:t:ˑi:ma	get.dressed:1PRS.PL	
		VC:ˑV:	kat:ˑi:ma	cover:1PRS.PL	

The correlations of the variables are shown in Table 17. The *long V2* level factor was chosen as a baseline and compared with *short V2* (**v2lengthshort**). The analysis shows that not all correlations are statistically significant. V2 length does not affect the V2 duration (t value = 0.345), while V1 length does affect it (t value = 6.198).

Another predictor that shows significant effect is the *sonorant* (**son**). The tokens with no sonorant were chosen as a baseline, and compared with those with a sonorant (**sonyes**); the analysis shows a negative correlation (t value = -5.386), which means that tokens with a sonorant (e. g. *hark:ˑa:ma* “step:1PRS.PL”) have a V2 shorter than tokens without a sonorant (e.g. *kat:ˑi:ma* “cover:1PRS.PL”).

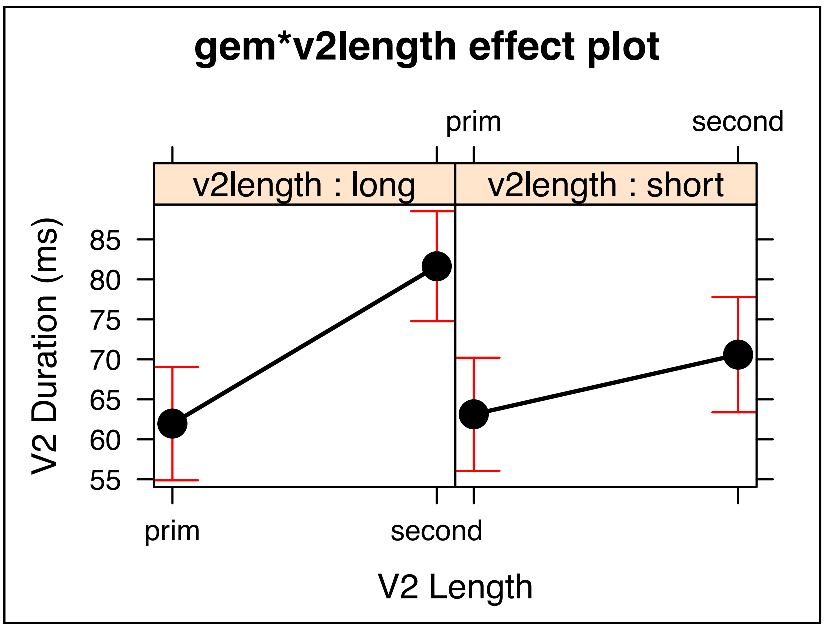
Table 17: Correlation of Variables, V2 Duration

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	-7.517	6.494	-1.157
gemsecond	19.673	3.334	5.901
v2lengthshort	1.164	3.369	0.345
v1lengthshort	**17.263**	**2.785**	**6.198**
sonyes	**-15.130**	**2.809**	**-5.386**
dur	0.097	0.007	14.140
gemsecond: v2lengthshort	**-12.216**	**4.708**	**-2.595**

The result of the interaction between C2 and V2 lengths (**gemsecond:v2lengthshort**) gives a statistically significant result (t value = -2.595). The base line is the *V2 duration* in a foot structure with a *primary geminate* and a *long V2* (V(:)**C**:ˑ**V:**), and it is compared with the *V2 duration* in foot structures with a *secondary geminate* and a *short V2* (V(:)**C**:**V**). There is no straightforward way to interpret the estimate (Baayen, 2008, p. 166), and although this correlation is significant, it is probably not perceptible (mean durations of short and long V2s after primary geminates differ by 15-16 ms); we will return to this in a subsequent analysis of a smaller subset, analyzing specifically the foot structures with secondary geminates.

Figure 9 plots the results of the correlation with the interaction between C2 duration and length of V2 with long vowels in the left, and short vowels in the right box. Each box is divided with respect to the geminate that precedes the vowel: primary on the left, and secondary on the right side of each box.

Figure 9. Durational Distribution of Second Syllable Vowels after Primary and Secondary Geminates



The mean duration of long and short vowels after primary geminates, and the short vowels after secondary geminates, does not significantly differ (about 62 ms). This means that the phonemic durational contrast between short and long vowels that follow primary geminates has been lost. The average duration of short and long V2s after primary geminates is not different. At the same time, there is no difference between the duration of the originally short V2 after secondary geminate, and the originally short V2 after primary geminates. This is statistical evidence that the structures with a secondary geminate followed by a short vowel were possible historically.

Table 18 shows the tested tokens with the new phonetic transcription, based on the test results.

Table 18: Shortening of Long V2 after Primary Geminates

V2 length	Foot Structure	Example	Ponetic transcription	Gloss
Long V2	VRC:ˑV:	hark:ˑa:ma	[hark:ˑ**a**ma]	step:1PRS.PL
	V:C:ˑV:	va:t:ˑi:ma	[va:t:ˑ**i**ma]	get.dressed:1PRS.PL
	VC:ˑV:	kat:ˑi:ma	[kat:ˑ**i**ma]	cover:1PRS.PL
Short V2	VC:ˑV	kat:ˑila	[kat:ˑ**i**la]	boiler:NOM
	V:C:ˑV	sa:t:ˑaja	[sa:t:ˑ**a**ja]	guide:NOM
	VRC:ˑV	tark:ˑoja	[tark:ˑ**o**ja]	smart:PART.PL

This result differs from that on disyllabic foot, where the vowels that follow primary geminates were different in duration (Markus, 2011). As I mentioned above, short vowels after secondary geminates do not occur in disyllabic feet. The distribution of the long vowels' duration is analyzed below.

Experiment 4. Interaction of V1 and V2 in structures with secondary geminates

In this experiment a multiple linear regression with interaction (mod 4, subset 3) was calculated to predict V2 duration based on **v1length, son,** and **dur**. The subset contains only tokens with a *secondary geminate* and a *long V2.* Structure types are presented in Table 19.

Table 19: Structures Tested: Long V2s after Secondary Geminates

V1 Length +son	Foot structure	Example	Gloss	N tokens
Short V1	VC:V:	mat:a:la	low:NOM	180
Long V1	V:C:V:	la:t:i:ma	get.ready:1PRS.PL	365
Short V1 + son	VRC:V:	kerk:i:mä	be.on.time:1PRS.PL	180
Long V1 + son	V:RC:V:	vä:nt:i:mä	turn.around:1PRS.PL	203

The statistical model estimation is provided in Table 20.

Table 20: Statistical Model Estimations

Predictor	AIC	p value
dip	-6	<0.0001
v1length	0	<0.0001
son	0	<0.0001
son:v1length	-22	<0.0001
dur	-247	<0.0001

Based on the results of the backward elimination test, the final model includes following predictors: **dip, dur,** and **son:v1length** interaction.

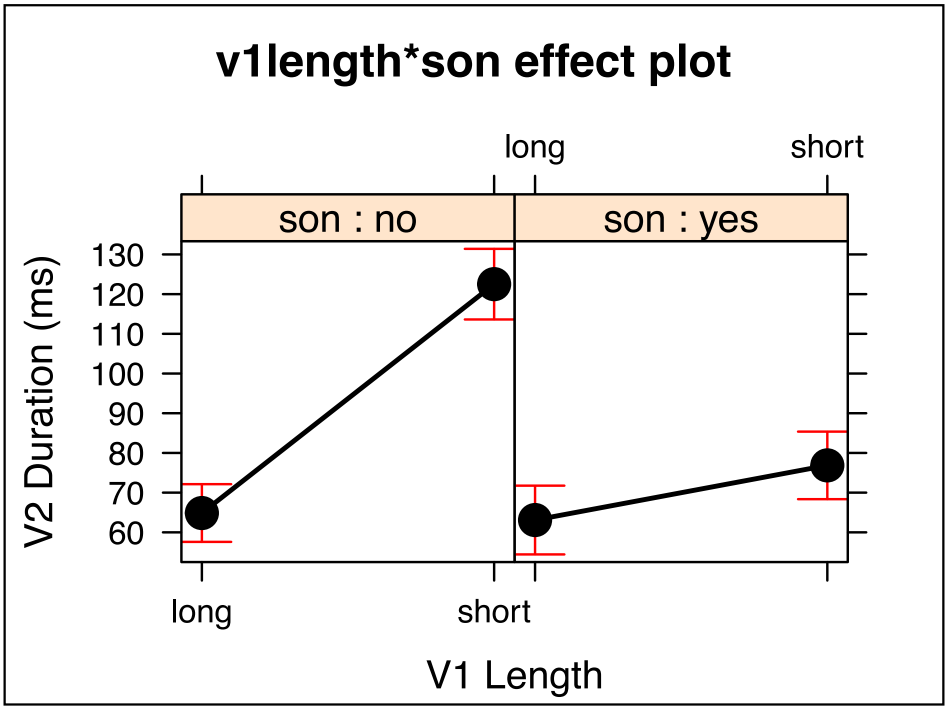
Table 21 shows the correlation effects. The duration of V2 in structures with a (**v1length-long:sonyes**) is significant (t value = 5.722).

Table 21: Correlation of Variables, V2 Duration

	Estimate	Std. Error	t value
(Intercept)	22.617	6.721	3.365
dipyes	14.118	5.334	2.647
v1lengthlong	-57.651	5.533	-10.420
sonyes	-45.651	5.425	-8.415
dur	0.140	0.008	17.538
v1lengthlong:sonyes	**43.885**	**7.669**	**5.722**

Figure 10 plots the results. Two boxes show the foot structures with a sonorant (son : yes) and without a sonorant (son : no); y-axis represents V2 duration in milliseconds; x-axis represents the V1 lengths (long and short) for each structure type.

Figure 10. Durational Distribution of Long and Short Vowels after Secondary Geminates



It can be seen that the only structure where V2 duration is substantially longer (120 ms against 60-78 ms), is those with no sonorant (son:no) *and* a short V1 (VC:V:). All other structures with the original long V2 have lost the durational contrast with those that originally had a short V2 (V:C:V:, VRC:V:, V:RC:V:). Table 22 shows the tested tokens with the new phonetic transcription, based on the test results.

Table 22: Shortening of Long V2 after Secondary Geminates

	Foot structure	Example	Phonetic Transcription	Gloss
Short V1	VC:V:	mat:a:la	[mat:**a**:la]	low:NOM
Long V1	V:C:V:	la:t:i:ma	[la:t:**i**ma]	get.ready:1PRS.PL
Short V1 + son	VRC:V:	kerk:i:mä	[kerk:**i**mä]	be.on.time:1PRS.PL
Long V1 + son	V:RC:V:	vä:nt:i:mä	[vä:nt:**i**mä]	turn.around: 1PRS. PL

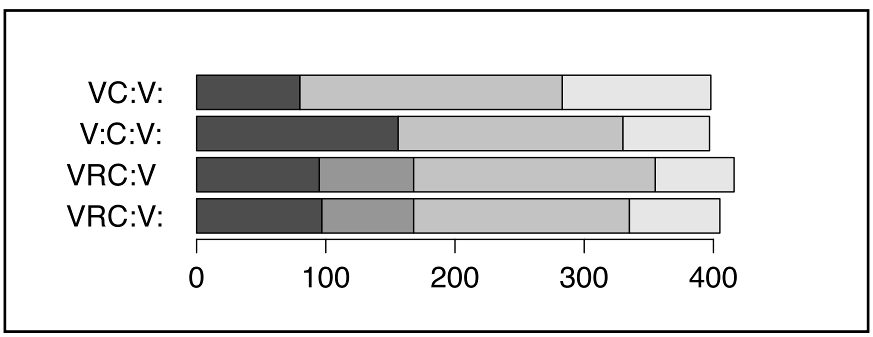
Summary

Summarizing the findings of the experiments 1–4, I can conclude the following:

1. Phonemic ternary contrast in consonants between first and second syllables has remained intact.
2. Phonemic binary contrast in V1s has remained intact.
3. Phonemic binary contrast in V2s that follow primary geminates has been lost. Original short V2s after secondary geminates are indeed short in duration.
4. Long and short V2s after secondary geminates are in complimentary distribution: long V2 occurs in structures with a short V1 and without a sonorant, while short V2 occurs elsewhere.

Figure 22 plots the average duration of the first syllable vowel (black), intervocalic consonant (gray) and second syllable vowel (light gray) in all possible structures with secondary geminates. The sonorant in last two structures is shown with dark gray.

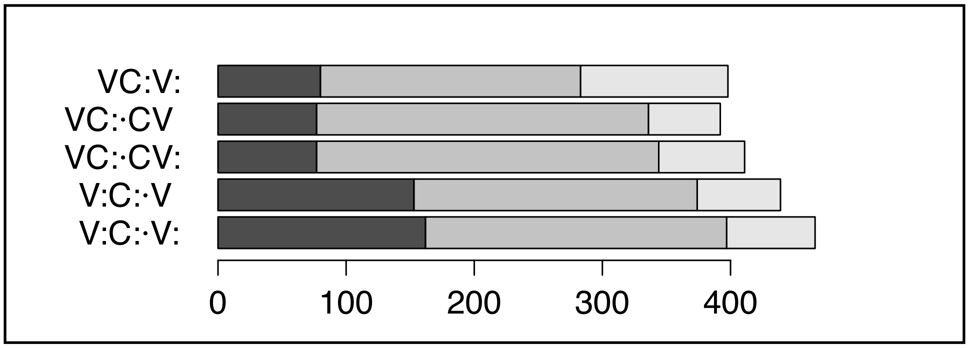
Figure 11: Foot Structures with Secondary Geminates: V2 Contrast



It can be seen that V2 is substantially longer only in the structure VC:V:. All other structures have short V2s, irrespective of the V2 original length.

Meanwhile, the contrast between V2s after primary geminates has been lost completely. Figure 12 plots the average durations of the nucleus elements in structures with primary geminates, and shows the lack of the difference between short and long V2s. I include a barplot of a structure with a long V2 after secondary geminate (VC:V:), so that the difference between short and long V2s is observable. It can be seen that in feet with primary geminates, there is no correlation between foot structure and V2 duration: all V2s are short.

Figure 12: Foot Structures with Primary Geminates: V2 Contrast Lost



6.2. Phonetic realization: Segment durational compensation

In this section I examine how elements of the trisyllabic foot nucleus influence each other's durations, testing if the phonemic length of one alternates the phonetic realization of the other. First I look at the interaction of the elements between first and the second syllables, and then test the V2 duration in the VCV: structure that is said to undergo phonetic lengthening in V2.

For two following tests, I use the same models as I did before, but run them on smaller datasets to see the correlations in detail.

Experiment 5: Influence of C2 length on Long V1 Duration

In this experiment a multiple linear regression (mod 5, subset 4) was calculated to predict long V1 duration based on **gem, son,** and **dur**. It is the same model I used to test V1 duration in the complete subset, run on a smaller subset of data to show more specific correlation, since the backward elimination test showed the lack of influence of C2 length on V1 duration (see Table 12).

Table 23 shows the structures selected for the test.

Table 23: Structures Tested: Long V1 Followed by Three Kinds of Consonants

C2 Length	Foot Structure	Example	Gloss	N tokens
Primary Geminate	V:C:ˑV	sa:t:ˑaja	guide:NOM	461
	V:C:ˑV:	va:t:ˑi:ma	get.dressed:1PRS.PL	
Secondary Geminate	V:C:V	na:t:ala	stove.bottom:NOM	698
	V:C:V:	la:t:i:ma	get.ready:1PRS.PL	
	V:RC:V	pi:nt:ara	seedbed:NOM	
	V:RC:V:	vä:nt:i:mä	turn.around: 1PRS.PL	141
Singleton	V:CV	sa:tama	accompany:1PRS.PL	
	V1V2CV	soitama	play:1PRS.PL	

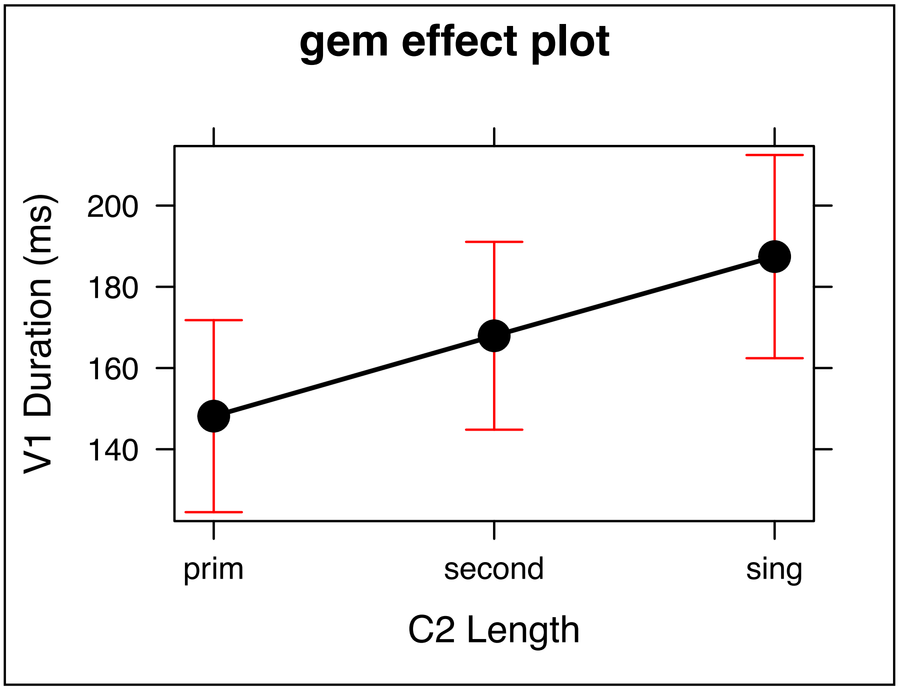
Table 24 shows the correlation effects. All effects are significant.

Table 24: Correlation of Variables, V1 Duration

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	16.073068	13.516955	1.189
gemsecond	**19.793699**	**5.377609**	**3.681**
gemsing	**39.294189**	**7.227560**	**5.437**
sonyes	-3.134164	5.806578	-0.540
dur	0.187165	0.008371	22.358

The V1 mean duration of structures with a primary geminate was chosen as a base line to compare with the durations of V1 in structures with other two types of consonants. Long V1 in structures with singletons (**gemsing**) are significantly longer (t value = **5.437**) then those in structures with primary geminates. The same effect can be seen with respect to secondary geminates (**gemsecond**) (t value = 3.681). Fig 11 plots the results. The y-axis represents the duration of a long V1 in ms, and x-axis corresponds with three types of consonants: primary and secondary geminates, and singletons.

Figure 11. Influence of C2 Length on V1 Duration



The perceptible difference, however, can be seen only in comparison of long V1 duration in structures with singletons vs. those in structures with primary geminates (estimate = 39 ms). In the case of structures with secondary geminates, although the difference between the mean durations is statistically significant, it is only 20 ms; this is below the perceptibility threshold, so it cannot be phonemic.

Experiment 6: The duration of vowels in feet with singletons

In this experiment, a multiple linear regression with interaction (mod 6, subset 5) was calculated to predict V2 duration based on **v1length** and **son** interaction**,** and **dur** in order to test, whether the V2 in VCV: foot structure undergoes prosodic lengthening. This subset contains only structures with singletons, which are listed in Table 25.

Table 25: Structures Tested: V2s after Singletons

sonorant	Nucleus structure	Example	Gloss	N tokens
	VCV:	ota:ma	take:1PRS.PL	67
	V:CV	sa:tama	accompany:1PRS.PL	98
+sonorant	VRCV	palkale	wage:ALL	36
	V:RCV	pi:ntara:	seedbed:PART:PL	43

Based on the results of the backward elimination test (Table 26), the final model includes following predictors: **dur,** and **son:v1length** interaction.

Table 26: Statistical Model Estimations

Predictor	AIC	p value
dip	+1	0.19
v1length	0	1
son	0	<0.0001
son:v1length	-18.7	<0.0001
dur	-12.1	<0.001

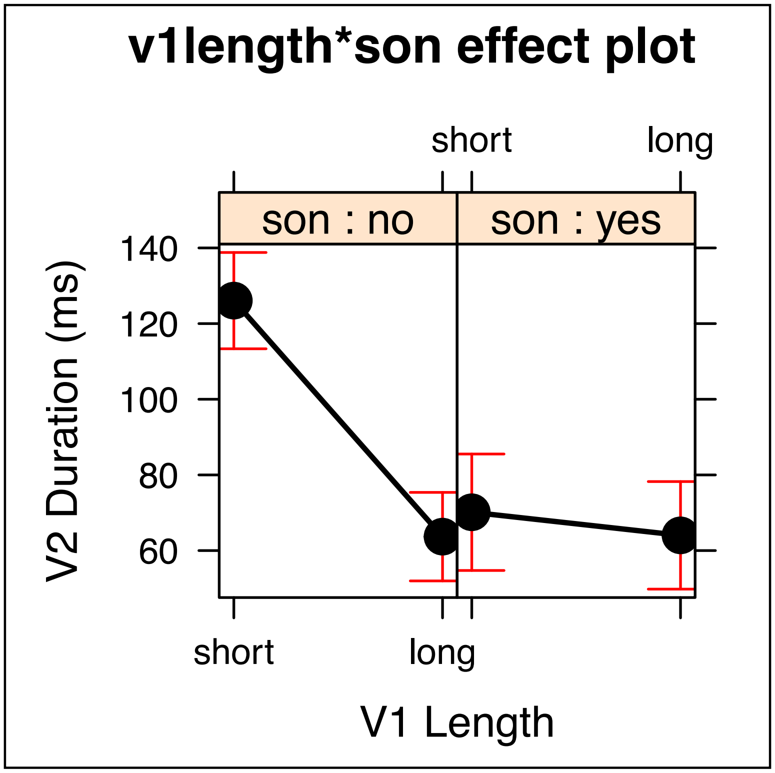
The base line is the *V2 duration* in a foot structure with a *short V1 and no sonorant* (V**CV:**), and it is compared with all other structures (V:CV, VRCV, V:RCV). All correlations were statistically significant (Table 27). The duration of vowels with no sonortant and short V1 differs in duration with those that have a sonorant (t value = -6.864), long V1 (t value = **-9.488**), and the interaction effect between **v1length** and **son** is significant as well (**5.213**).

Table 27: Correlation of Variables, V2 Duration

Fixed effects:	Estimate	Std. Error	t value
(Intercept)	86.3486	10.9773	7.866
v1lengthlong	**-62.3911**	**6.5759**	**-9.488**
sonyes	**-55.9429**	**8.1505**	**-6.864**
dur	0.0606	0.0156	3.885
v1lengthlong:sonyes	**56.2848**	**10.7963**	**5.213**

Figure 12 plots the results of the statistical analysis.

Figure 12. Durational Distribution of Long and Short Vowels after Singletons



The plot shows that the mean duration of V2 after singletons in the VCV: foot structure (120 ms) is significantly longer than the duration of V2 after all other structures (60-67 ms).

Summary

Summarizing, the results of this section, I conclude the following.

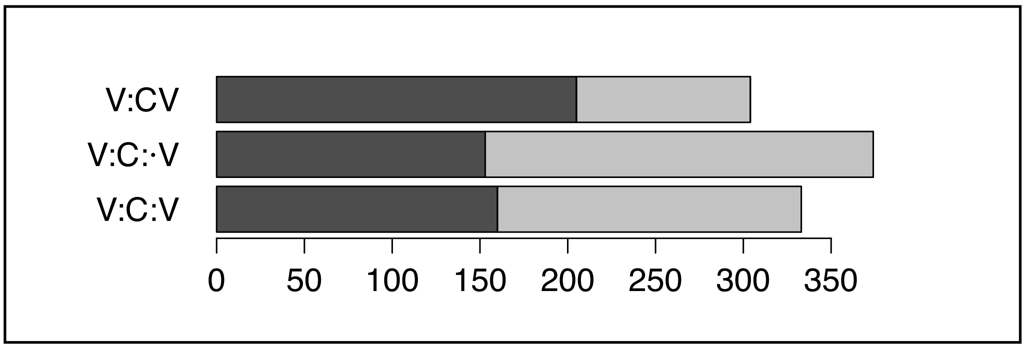
(1) In a trisyllabic foot, the duration of a first vowel and the following consonant have inverse relationship: a longer vowel is followed by a shorter consonant, and a shorter vowel is followed by a longer consonant. This results in three types of vowels on the phonetic level:

- short V1
- long V1 followed by a singleton
- long V1 followed by both primary and secondary geminates.

However, this difference between vowels that precede geminates, is statistically significant only between the following combinations: Long V1 + singleton vs. Long V1 + primary geminate.

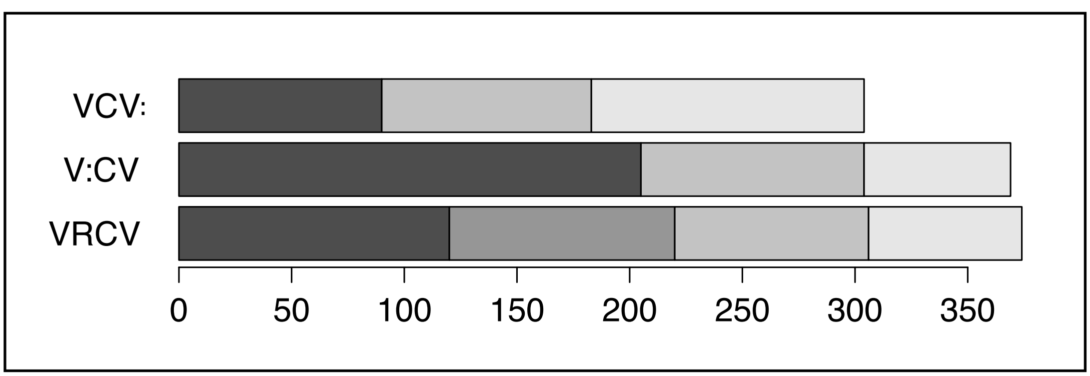
Figure 13 plots the mean durations of V1s and C2s. Black bars represent V1 durations, and gray bars represent C2 durations. V1 is clearly longer if followed by a singleton than if it is followed by geminetes.

Figure 13: Difference in V1 Duration Influenced by C2 Length



(2) In feet with singletons, vowels in the second syllable undergo a prosodic lengthening in structures with a light first syllable (VCV:). Figure 14 plots the mean durations of three distinctive types of the singleton structures.

Figure 14: Difference in V2 Duration Influenced by C2 Length



7. Interpretation

The results show that vowels in the second syllable of trisyllabic words tend toward the reduciton and loss of the phonemic durational contrast. In disyllabic words the phonemic durational contrast in V2 was attested in all structures with an intervocalic primary geminate; long V2s were attested in all structures with the intervocalic secondary geminate (Kuznetsova 2012, Markus, 2011). In trisyllables long V2 was attested in a structure, where the first vowel is short (VC:V:), and in a structure where first syllable is short (VCV:). This is a result of the reduction process that Ingrian is undergoing (Markus 2010, 2011, Kuznetsova 2012, 2015, Kuznetsova at al. 2015), and the isochronic tendency.

More specifically, the results demonstrate that original phonemically long and short V2s do not differ in duration, and the long vowels are equal to short ones. Table 28 lists all foot nucleus structures with primary geminates (structures with reduced long V: in the second syllable are shaded):

Table 28: Trisyllabic Foot Structures with Primary Geminates

Original V2 Length	Foot Nucleus	Example	Gloss	Phonetic Transcription
Short	VC:ˑV	kat:ˑila	boiler:NOM	[kat:ˑila]
	V:C:ˑV	sa:t:ˑaja	guide:NOM	[saat:ˑaja]
	VRC:ˑV	tark:ˑoja	smart:PART.PL	[tark:ˑoja]
Long	VC:ˑV:	kat:ˑi:ma	cover:1PL	[kat:ˑ**i**ma]
	V:C:ˑV:	vaat:ˑi:ma	dress:1PL	[vaat:ˑ**i**ma]
	VRC:ˑV:	hark:ˑa:ma	step:1PRS.PL	[hark:ˑama]

In feet with secondary geminates, the only structure that has preserved the long V2 is VC:V: — the foot with a light fist syllable (a short V1 and no sonorant). This effect is summarized in Table 29, structures with the reduced long V2 are shaded.

Table 29: Trisyllabic Foot Structures with Secondary Geminates

V2 Original Length	Foot Nucleus	Example	Gloss	Phonetic Transcription
Long	VC:V:	pap:e:ri	paper:NOM	[pap:**e:**ri]
	V:C:V:	la:t:i:ma	get.ready:1PRS.PL	[la:t:**i**ma]
	VRC:V:	hark:a:ma	step:1PRS.PL	[hark:**a**ma]
	V:RC:V:	vä:nt:i:mä	turn.around: 1PRS.PL	[vä:n:**i**mä]

At the same time, the under-investigated structure with a short V2 was attested, and this is consistent with the evidence provided in older research, that those structures existed in Soikkola Ingrian. Table 30 lists those structures with the phonetic transcription based on the results of the conducted analysis.

Table 30: Trisyllabic Foot Structures with Secondary Geminates

V2 Original Length	Foot Nucleus	Example	Gloss	Phonetic Transcription
Short	V:C:V	na:t:ala	stove.borrom:NOM	[na:t:**a**la]
	VRC:V	märk:eni	rot:3SG.PST	[märk:**e**ni]
	V:RC:V	pi:nt:ara	seedbed:NOM	[pi:nt:**a**ra]

At the same time, results show that long and short V2s after secondary geminates are in the complimentary distribution. Long V2 occurs in structures with a short V1, while short V2 occurs elsewhere. Therefore, the phonological status of V2 length has been lost.

As for structures with singletons, results are consistent with the system attested in disyllables: V2s in feet with light first syllable undergo prosodic lengthening, while in all other structures they remain phonetically short.

Thus, the number of distinctive phonemic structures with primary and secondary geminates from 22, as I had initially, has reduced to 11, listed in Table 31.

Table 31: All Trisyllabic Foot Structures in Contamporary Soikkola Ingrian

1	Primary Geminates	VC:·V	kat:·ila	boiler:NOM
2		V:C:·V	sa:t:·aja	guide:NOM
3		VRC:·V	tark:·oja	smart:PART.PL
4	Secondary Geminates	VC:V:	pap:e:ri	paper:NOM
5		V:C:V	na:t:ala	stove.borrom:NOM
6		VRC:V	märk:eni	rot:3SG.PST
7		V:RC:V	pi:nt:·ara	seedbed:NOM
8	Singletons	VCV:	ota:ma	take:1PRS.PL
9		V:CV	sa:tama	accompany:1PRS.PL
10		VRCV	palkale	wage:ALL
11		V:RCV	pi:ntara:	seedbed:PART:PL

Based on this analysis, I can formulate the current language tendency as follows: a foot in Soikkola Ingrian is constrained by a minimal and maximal cumulative nucleus duration. If a given nucleus is shorter than the “minimally allowed”, V2 undergoes a prosodic lengthening (*VCV > VC**V**:); if it is longer then the “maximally allowed”, the second vowel gets reduced (*VC:·**V**: > V:C:·**V**, *V:C:·**V**: > V:C:·**V**, *VRC:·**V**: > VRC:·**V**). Thus, a primary geminate makes the foot nucleus too long, so the V2 is reduced in all structures that contain a primary geminate. Same applies to the words with a secondary geminate preceded by a long vowel, or a sonorant, or both (*V:C:**V**: > V:C:**V**, *VRC:**V**: > VRC:**V**, *V:RC:**V**: > V:RC:**V**).

Only syllables with the short first vowel and either a singleton or a short geminate (VCV:, VC:V:) allow the V2 to be long. However, those V2s are long for different reasons: in the VCV: structure, the V2 is lengthened, because the cumulative foot duration is shorter than the possible minimum. In the VC:V: foot structure, V2 was originally long; it was not reduced, since it did not exceed the durational foot limit. Finally, if a word fits the given time span, V2 stays intact (VC:V:). In all foot types with primary geminates the V2 is short, since those structures exceed the maximal possible foot duration. In disyllables the cumulative foot duration fits the “allowed” timespan.

The Soikkola Ingiran foot structurally varies more than Estonian, because the segments are not as tightly bound, since short and long vowels are still contrastive in non-initial syllables, at least in disyllables. According to results provided by Markus (2011), in the disyllabic foot the contrast between V2s after primary geminates has remained.

On the other hand, the Soikkola Ingrian foot yet differs from the Finnish prosody system, due to the severe reduction process that in the first place affects the V2 length. The Votic language, spoken in the same area, used to have the northern type of contrast in non-initial vowels (*long vs. short*), but it was recently transformed to *full vowel vs. reduced vowel* in Luutsa Votic, and *vowel vs. no vowel* in Jõgõperä Votic (Markus 2011). Soikkola Ingrian seems to follow the same tendency. We can observe the reduction process in Soikkola Ingrian in progress: long duration which still can be found in shorter structures is getting reduced in longer structures.

The effect of the experimental settings on the results. In this experiment, all the tokens were pronounced sentence finally, and it is likely that the durational contrasts were strengthened by the experiment setting itself. The prosodic peak of a sentence accelerates the lengthening effect, at least in English (White, 2010, p. 460). Since all V1s are stressed vowels, and pronounced sentence finally with a prosodic peak on the last word, this might reinforce the stress effect that is manifested in duration. According to White (2014), final lengthening as a result of the prosodic peak can affect the final stressed vowel which is the most susceptible element to prosodic final lengthening. This methodological problem makes the interpretation more difficult.

It is also likely that in the spontaneous speech the results will differ, and some of the contrasts will be reduced even more. Therefore, in order achieve an exhaustive description of the contemporary Soikkola Ingrian prosodic system, more experiments utilizing different conditions are needed, including the design of questionnaires, and the elicitation methods.

The next step of this study would be the analysis of structures that remain under-investigated, such as disyllabic feet with a closed second syllable and with clusters between first and second syllables, monosyllabic feet; words that have a geminate between second and third syllbles. Another path this study could take is the investigation of the vowel quality; comparison of reduced long with original short vowels will help describe the system of phonological contrasts in Ingrian more accurately.

8. Conclusions

In this study I aimed to investigate which durational contrasts remain in the trisyllabic foot of the Soikkola Ingrian dialect, and test how those contrasts are realized on the phonetic level. I was specifically interested if the ternary contrast in consonants between first and second syllables, and binary contrast in first and second syllable vowels. These contrasts have remained. Next, I was testing if structures with short vowels that follow secondary geminates were possible; the evidence about this type of foot structure was found in old studies, but this contradicted the historical presupposition of the secondary gemination conditions.

The study showed that except for the vowels in the second syllable, the Soikkola Ingrian dialect has the same durational contrasts in the trisyllabic foot as it has in the disyllabic one:

- Consonants between first and second syllables show the ternary durational opposition.
- First syllable vowels show the binary durational opposition.

However, with respect to the vowels of the second syllable, the trisyllabic foot shows discrepancies with the results of the disyllabic foot analysis:

- The durational contrast of long vs. short V2s has completely disappeared after primary (long) geminates (unlike disyllables).
- In the V2s that follow secondary (short) geminates, long vowels were found only in VC:V: structure; short V2 is found elsewhere (V:C:V:, VRC:V:, V:RC:V:).
- This means that V2 length is not phonemic and; long and short V2s are distributed complimentary.
- In all structures with the original short V2 (V:C:V, VRC:V, V:RC:V) V2 duration is indeed short.

On the phonetic level, the shortening and the lengthening of foot nucleus elements follow the isochronic tendency that was attested in disyllables in previous research:

- V1 is significantly shorter before primary geminates than before singletons
- Just like in disyllables, short V2 in the VCV: foot nucleus, the V2 undergoes a phonetic lengthening.

The results show that Soikkola Ingiran is highly susceptible to the isochrony and reduction effects. It undergoes a prosodic change from the “northern”, type of Finnic languages, where the duration of the second vowel is contrastive (similar to Finnish) to the “southern” type, where the duration of the second syllable vowel depends on the structure of the first syllable (similar to Estonian). The duration of segments highly depends on the foot structure.

9. Bibliography

1. Acta Linguistica Petropolitana (2012). vol. 8, part 2. S-Pb.: Nauka, 2012. 1024 p.
2. Aoyama, K. (2001). *A psycholinguistic perspective on Finnish and Japanese prosody: Perception, production and child acquisition of consonantal quantity distinctions*, 3916, Springer Science & Business Media.
3. Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, 40(3), 351–373.
4. Asu, E. L. (2004). *The Phonetics and Phonology of Estonian Intonation*. Doctoral dissertation. University of Cambridge
5. Bates D, Maechler M, Bolker B and Walker S (2014). _lme4: Linear mixed-effects models using Eigen and S4_. R package version 1.1-7, <URL: http://CRAN.R-project.org/package=lme4>
6. Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390–412.
7. Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.
8. Boersma, Paul (2001). Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341-345.
9. Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of phonetics*, 29(2), 217–230.
10. Gordon, M. (2009). Prominence and gemination in Ingrian. *Linguistica Uralica*, (2), 81–100.
11. Hogan, J. T., & Rozsypal, A. J. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *The Journal of the Acoustical Society of America,* 67(5), 1764–1771.
12. Johnson, K. (2011). Quantitative methods in linguistics. John Wiley & Sons.
13. Kazanskiy N. (2012). Appendix 7. Rules of Phonetic and Phonological transcription. *Acta Linguistica Pertopolitana*. Saint-Petersburg, 2012. (In Russian)
14. Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5), 1208–1221.
15. Krull, D. (1999). Foot isochrony in Estonian. *Proceedings of the 14th International Congress of Phonetic Sciences*. San Francisco, 3, pages 1063–1066.

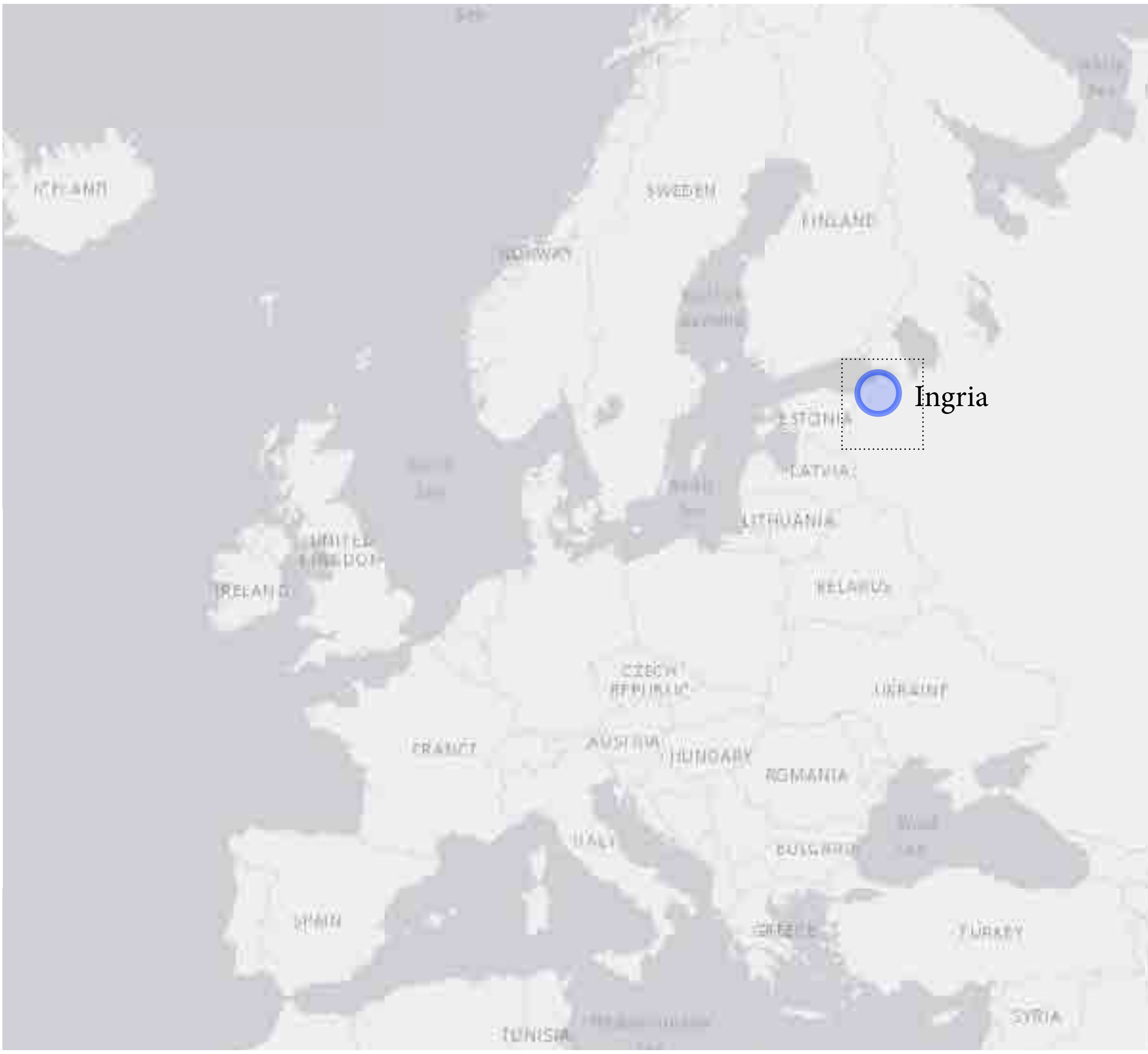
16. Kuznetsova N. (2009a) Fonologicheskie sistemy izhorskih dialektov. Dissertatsiya na soiskanit uchenoy stepeni kandidata fililogicheskih nauk. Sankt-Peterburg (Phonological systems of Ingrian dialects. Unpublished PhD thesis.)
17. Kuznetsova N. (2009b) Suprasegmentnaya fonologiya soikinskogo dialekta izhorskogo yazyka v tipologicheskom aspekte. *Voprosy Yazykoznaniya,* 5, p. 18–47.
18. Kuznetsova N. (2012) Finnic foot nucleus lengthening: From phonetics to phonology. *Nordic prosody: Proceedings of the XIth Conference*. Tartu 2012. p. 205–214
19. Kuznetsova N. (2015a) Evolution of the non-initial vocalic length contrast across Finnic varieties of Ingria and the adjacent areas. *Linguistica Uralica*, 51(4), 2015.
20. Kuznetsova, N. (2015b). Two phonological rarities in Ingrian dialects. *New Trends in Nordic and General Linguistics*, 42, 91.
21. Kuznetsova, N., Markus, E., & Muslimov, M. (2015). Finnic Minorities of Ingria. *Cultural and Linguistic Minorities in the Russian Federation and the European Union: Comparative Studies on Equality and Diversity.* Springer International Publishing, 127–167.
22. Knight, R. A. (2011). Assessing the temporal reliability of rhythm metrics. *Journal of the International Phonetic Association*, 41(03), 271–281.
23. Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages.* Oxford & Cambridge, MA: Blackwell Publishers Inc., 1996
24. Laanest A. (1966). *Izhorskie dialekty. Lingvogeograficheskoe issledovanie.* Tallinn: Institut jazyka i literatury AN SSSR.
25. Laanest A. (1978). *Istoricheskaya fonetika i morfologiya izhorskog yazyka.* Diss. kand. filol. nauk. Tallinn: Institut jazyka i literatury AN SSSR.
26. Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. *American studies in Uralic linguistics*, 1, 21–82.
27. Lehiste, I. (1965). The Function of Quantity in Finnsh and Estonian. — *Language* 41, 447–456.
28. Lehiste, I. (1977). Isochrony reconsidered. *Journal of phonetics*, 5, 253–263.
29. Lehiste, I. (2008). Lehiste, Ilse, Pire Teras, Valts Ernštreits, Pärtel Lippus, Karl Pajusalu, Tuuli Tuisk, Tiit-Rein Viitso (2008) *Livonian prosody.* (MSFOu, vol. 255) Suomalais-ugrilainen seura.
30. Lehtonen J. (1970). *Aspects of quantity in standard Finnish*. Jyväskylä: University of Jyväskylä.
31. Markus, E. (2010). Primary and secondary geminates in Ingrian. *Linguistica Uralica*, (1), 38–52.

32. Markus, E. (2011). The phonetics and phonology of a disyllabic foot in Soikkola Ingrian. *Linguistica Uralica*, (2), 103-119.
33. Markus, E. (in press). E. Markus, P. Lippus, K. Pajusalu, and P. Teras. http://www.researchgate.net/publication/259900380_Three-way_opposition_of_consonant_quantity_in_Finnic_and_Saamic_languages. Retrieved on: October 29, 2015
34. Nolan, F., & Jeon, H. S. (2014). Speech rhythm: a metaphor? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 369(1658).
35. Oostendorp, M., van (2005). Mora Theory. Handout of talks presented at Universiteit Leiden, the Netherlands. http://www.researchgate.net/profile/Marc_Oostendorp/publication. Retrieved on: April 29, 2015
36. Porkka, V. (1885). *Über den ingrischen Dialekt: mit Berücksichtigung der übrigen finnisch-ingermanländischen Dialekte*, Helsinki. Helsingfors: J. C. Frenckell & Sohn, 1885. – 147 p.
37. Prillop, K. (2013). Feet, syllables, moras and the Estonian quantity system. *Linguistica Uralica*, (1), 1-29.
38. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/
39. Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *Proceedings of the 15th International Congress of Phonetic Sciences,* ICPhS 15-Barcelona, Spain, 337-342.
40. Sovijärvi, A. (1944). *Foneetis-äännehistoriallinen tutkimus soikkolan inkeroismurteesta*, Helsinki, Suomalaisen kir jallisuuden seura.
41. Topintzi, N. (in press). *Onset, stress and weight interactions.* Unpublished Ph.D. thesis, University College London.
42. Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, 28(4), 397-440.
43. Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, 4(1), 93-118.
44. Viitso T.-R. (1961) *Vadja Luutsa-Liivčülä murraku fonoloogia*. Emakeele seltsi aastaraamat 1961, VII. Lk. 142–174.
45. White, L., & Turk, A. E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, 38(3), 459-471.

46. White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication*, 63, 38-54.
47. Wiik, K. (1991) On a third type of speech rhythm: foot timing. *Proceedings of the XIIth International Congress of Phonetic Sciences, Aix-en-Provence, August 19-24*. Vol 3, 298–301.

APPENDIX I.

A. Map of Europe[1]



B. Soikkola Peninsula

1 Maps are generated at: mapmaker.education.nationalgeographic.com

APPENDIX II. Words Analyzed

Wordlist A: Trisyllables with Primary and Secondary Geminates

C:ˑV	C:ˑV:	C:V	C:V:
VC(:)C			
kat:ˑila boiler:NOM	kat:ˑi:ma cover:1PL		mat:a:la low:NOM
pet:ˑeri Peterburg:NOM	mit:ˑa:ma try.on:1PL		put:e:li bottle:NOM
kuk:ˑoro wallet:NOM	lük:ˑä:mä push:1PL		väk:ö:vä prolific:NOM
kak:ˑara pancake:NOM	huk:ˑa:ma lose:1PL		pik:a:ri winglass:NOM
(elä) tap:ˑele fight:CNG	hüp:ˑi:mä jump:1PL		pap:e:ri paper:NOM
V:C:(ˑ)			
sa:t:ˑaja guide:NOM	va:t:ˑi:ma get.dressed:1PL	na:t:ala stove.bottom:NOM laatana incense:1PL	la:ti:ma get.ready:1PL
u:t:ˑele wait:2SG	mu:t:ˑiima change:1PL	vu:t:ava leaking:NOM	su:ti:ma judge:1PL la:ti:ma get.ready:1PL
pli:t:ˑoja stove:PART:PL	sü:t:ˑiimä feed.oneself:1PL	(elä) ri:t:ele quarrel:CNG	pre:ti:mä talk.deliriously:1PL
mi:k:ˑula Nikolas:NOM	muvk:ˑaama torture:1PL	pi:k:oja servant:1PL:PART	hu:ka:ma rest:1PL
pu:k:ˑoja switchblade:PART	pra:k:ˑaama reject:1PL	si:k:oja cisco:1PL:PART	
V1V2C:(ˑ)			
voit:ˑeli fight:3PST	toit:ˑi:ma feed.oneself:1PL	voit:eli smear:3SG:PST	hoit:i:ma beware:1PL

käüt:ˑeli lead:3PST	näüt:ˑi:mä like:1PL	laut:oja plank:1PL:PART	haut:a:ma tomb:1PL
soik:ˑola Soikkola:NOM (a peninsula name)	luik:ˑa:ma slurp:1PL	oik:eni right. oneself:3SG:PST	oik:a:ma smooth.down:1PL
lauk:ˑoja onion:PL:PART Loukkula Loukula:NOM (a village name)	hauk:ˑa:ma bite:1PL	peuk:olo thumb:NOM	lauk:a:ma destroy:1PL
	louk:ˑa:ma press:1PL	kiuk:ata stove:PART	hiuk:a:ma hiccup:1PL
VRC:(ˑ)			
tark:ˑoja smart:PL:PART	törk:ˑä:mä rub:1PL	märk:eni rotten:3PST	kerk:i:mä have.time:1PL
mark:ˑoja stamp:PL:PART	hark:ˑa:ma step:1PL	murk:ina breakfast:NOM	hülk:ä:mä undress:1PL
palk:ˑoja salary:PL:PART	palk:ˑa:ma hire:1PL	jalkvoja leg:1PL:PART	hülk:ä:mä раздеваемся 1PL
kelk:ˑoja sleigh:PART	solk:ˑa:ma flick:1PL	velk:oja debt:PART:PL	kert:a:ma twist:1PL
kert:ˑele touch:CNG	sort:ˑu:ma sort:1PL	mertvoja basket:PART:PL	
V:RC			
		pi:nt:ara seedbed:NOM	ki:lt:i:mä decline:1PL
		vä:nt:eli toss:3SG:PST	vä:nt:ivmä toss:3SG:1PL
		kä:nt:eli roll:3SG:PST	kä:nti:mä roll:3SG:1PL
		ki:rt:eli turn:3SG:PST	ki:rti:mä turn:3SG:1PL
		vi:rt:eli spin:3SG:PST	vi:rti:mä spin:3SG:1PL

Wordlist B: Trisyllables with Singletons

VCV	otama	take:1PL
	jätämä	leave:1PL
	ukole	old.man:ALL
	kukale	flower:ALL
	akale	old.woman:ALL
	tapama	kill:1PL
VVCV	sa:tma	accompain:1PL
	su:tuma	be.mad:1PL
	sü:tämä	feed:1PL
V1V2CV	soitama	play:1PL
	näütämä	show:1PL
	koikale paikale	bed:ALL glew:1PL
	laukale	onions:ALL
	haukata	bite:INF
	loukata	squeeze:SUP
VRCV	törkätä	grate:SUP
	kirkole	church:INF
	palkale	salary:ALL
	kelkale	sleigh:ALL
V1V2RCV	pi:ndara:	seed.bed:PART:PL
	vä:ndelö:	turn:3SG
	kä:ndelö:	tumble:3SG
	ki:rdelö:	spin:3SG
	vi:rdelö:	roll:3SG

APPENDIX III: Linear Regression Models

Mod #	Dependent Variable	# of Speakers	Predictors	Data Subset
1	C2 Duration	3	Consonant Type Foot Duration Sonorant	1
2	V1 Duration	3	Consonant Type Foot Duration Sonorant	1
3	V2 Duration	5	Consonant Type*V2 length Foot Duration Sonorant	2
4	V2 Duration	5	V1 length*son Foot Duration	3
5	V2 Duration	5	Consonant Type Foot Duration Sonorant	4
6	C2 Duration	3	Consonant Type*V1 Length Foot Duration Sonorant	5

mod1
Linear mixed model fit by REML ['lmerMod']
Formula: c2.dur ~ son + gem + v1length + dur + v2length + (1 | word) + (1 | spkr)
Data: d_ms_s

REML criterion at convergence: 20562

Scaled residuals:

Min	1Q	Median	3Q	Max
-4.4813	-0.6086	-0.0540	0.5699	6.4776

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	188.7	13.74
spkr	(Intercept)	495.7	22.26
Residual		617.7	24.85
Number of obs: 2197, groups:	word, 133; spkr, 3		

Fixed effects:

	Estimate Std.	Error	t value
(Intercept)	119.13126	16.35581	7.284
sonyes	-32.09235	3.25524	-9.859
gemsecond	-51.66617	3.12670	-16.524

gemsing	-130.46988	4.92272	-26.504
v1lengthshort	27.38906	3.31262	8.268
dur	0.13971	0.00668	20.913
v2lengthlong	12.10865	8.76163	1.382
v2lengthshort	18.10622	8.39552	2.157

Correlation of Fixed Effects:

(Intr)	sonyes	gmscnd	gemsng	v1lngt	dur	v2lngthl	
sonyes	0.166						
gemsecond	-0.199	-0.298					
gemsing	-0.353	-0.196	0.363				
v1lngthshrt	-0.295	-0.468	0.309	0.222			
dur	-0.266	-0.052	0.117	0.029	0.111		
v2lengthlng	-0.499 -	0.276	0.105	0.561	0.323	-0.095	
v2lngthshrt	-0.515	-0.323	0.140	0.497	0.388	-0.052	0.941

mod2

Linear mixed model fit by REML ['lmerMod']

Formula: v1.dur ~ v1length + gem + dur + (1 | word) + (1 | spkr)

Data: d_ms_s

REML criterion at convergence: 20347.8

Scaled residuals:

Min	1Q	Median	3Q	Max
-4.4741	-0.5714	-0.0174	0.5575	5.0720

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	274.7	16.57
spkr	(Intercept)	169.2	13.01
	Residual	542.8	23.30

Number of obs: 2197, groups: word, 133; spkr, 3

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	32.068281	9.329562	3.437
v1lengthshort	-64.642168	3.231873	-20.001
gemsecond	15.091219	3.449248	4.375
gemsing	35.130800	4.646574	7.561
dur	0.167948	0.006265	26.808

Correlation of Fixed Effects:

	(Intr)	v1lngt	gmscnd	gemsng
v1lngthshrt	-0.219			
gemsecond	-0.258	0.191		
gemsing	-0.191	0.058	0.384	
dur	-0.505	0.094	0.087	0.092

mod3

Linear mixed model fit by REML ['lmerMod']
Formula: v2.dur ~ gem * v2length + v1length + son + dur + (1 | word) +
(1 | spkr)
Data: d_ms

REML criterion at convergence: 30450.1

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.8941	-0.3466	-0.0398	0.2701	27.6179

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	110.76	10.524
spkr	(Intercept)	36.11	6.009
	Residual	983.69	31.364

Number of obs: 3115, groups: word, 123; spkr, 5

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-7.516806	6.494311	-1.157
gemsecond	19.673253	3.334032	5.901
v2lengthshort	1.163661	3.369376	0.345
v1lengthshort	17.263053	2.785294	6.198
sonyes	-15.130311	2.809390	-5.386
dur	0.097047	0.006863	14.140
gemsecond:v2lengthshort	-12.216057	4.707745	-2.595

Correlation of Fixed Effects:

	(Intr)	gmscnd	v2lngt	v1lngt	sonyes	dur
gemsecond	-0.389					
v2lngthshrt	-0.369	0.513				
v1lngthshrt	-0.255	0.151	0.022			
sonyes	0.040	-0.173	-0.014	-0.465		

dur	-0.815	0.155	0.144	0.123	-0.068	
gmscnd:v2ln	0.206	-0.658	-0.708	0.121	-0.078	-0.054

mod4

Linear mixed model fit by REML ['lmerMod']
Formula: v2.dur ~ v1length * son + dur + (1 | word) + (1 | spkr)
Data: v2long_second

REML criterion at convergence: 8037.5

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.1021	-0.6260	-0.0423	0.5641	4.9886

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	87.21	9.338
spkr	(Intercept)	18.19	4.265
Residual	317.41	17.816	

Number of obs: 928, groups: word, 34; spkr, 5

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-27.669070	6.619602	-4.180
v1lengthshort	50.387066	5.075133	9.928
sonyes	-8.937844	4.950867	-1.805
dur	0.139787	0.008006	17.460
v1lengthshort:sonyes	-36.590052	7.597282	-4.816

Correlation of Fixed Effects:

	(Intr)	v1lngt	sonyes	dur
v1lngthshrt -0.331				
sonyes	-0.239	0.315		
dur	-0.858	0.110	-0.003	
v1lngthshr:	0.188	-0.664	-0.651	-0.035

mod5

Linear mixed model fit by REML ['lmerMod']
Formula: v2.dur ~ dip + v1length * son + dur + (1 | word) + (1 | spkr)
Data: v2long_second

REML criterion at convergence: 8025.5

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.1144	-0.6502	-0.0473	0.5797	4.9938

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	75.46	8.687
spkr	(Intercept)	18.11	4.255
Residual	316.54	17.792	

Number of obs: 928, groups: word, 34; spkr, 5

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	-35.033975	7.151108	-4.899
dipyes	14.117662	5.333518	2.647
v1lengthshort	57.650681	5.532533	10.420
sonyes	-1.766752	5.423260	-0.326
dur	0.140054	0.007986	17.538
v1lengthshort:sonyes	-43.884547	7.669118	-5.722

Correlation of Fixed Effects:

	(Intr)	dipyes	v1lngt	sonyes	dur
dipyes	-0.405				
v1lngthshrt -0.462	0.504				
sonyes	-0.388	0.513	0.493		
dur	-0.800	0.021	0.111	0.008	
v1lngthshr:	0.30	-0.363	-0.718	-0.707	-0.042

mod6

Linear mixed model fit by REML ['lmerMod']
Formula: v1.dur ~ gem + son + dur + (1 | word) + (1 | spkr)
Data: d_ms_s_v1long

REML criterion at convergence: 12237.3

Scaled residuals:

Min	1Q	Median	3Q	Max
-4.5511	-0.5819	-0.0194	0.5510	4.8302

Random effects:

Groups	Name	Variance	Std.Dev.
word	(Intercept)	345.7	18.59
spkr	(Intercept)	385.7	19.64
Residual		631.4	25.13

Number of obs: 1300, groups: word, 80; spkr, 3

Fixed effects:

	Estimate	Std. Error	t value
(Intercept) 16.073068	13.516955	1.189	
gemsecond	19.793699	5.377609	3.681
gemsing	39.294189	7.227560	5.437
sonyes	-3.134164	5.806578	-0.540
dur	0.187165	0.008371	22.358

Correlation of Fixed Effects:

(Intr)	gmscnd	gemsng	sonyes	
gemsecond -0.241				
gemsing	-0.170	0.483		
sonyes	-0.001	-0.379	-0.267	
dur	-0.465	0.083	0.043	0.002

Linear mixed model fit by REML [‘lmerMod’]
Formula: v2.dur ~ v1length * son + dur + (1 | word) + (1 | spkr)
Data: sing_only