Journal of Phonetics 100 (2023) 101246

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Compensatory effects of foot structure in segmental durations of Soikkola Ingrian disyllables and trisyllables

Natalia Kuznetsova [a,b], Irina Brodskaya [c], Elena Markus [d,e]

[a] Università Cattolica del Sacro Cuore
[b] Institute for Linguistic Studies, Russian Academy of Sciences
[c] Queen Margaret University
[d] University of Tartu
[e] Institute of Linguistics, Russian Academy of Sciences

ARTICLE INFO

Article history:
Received 30 March 2022
Received in revised form 26 April 2023
Accepted 30 April 2023

Keywords:
Soikkola Ingrian (Finnic)
Trisyllabic foot
Disyllabic foot
Ternary quantity of consonants
Temporal compensation
Vowel shortening
Isochrony

ABSTRACT

This acoustic study explores compensatory influences of foot structure on segmental duration and quantity in the foot nuclei of 22 trisyllabic and four disyllabic structures in vanishing Soikkola Ingrian (Finnic). A robust ternary quantity contrast of consonants is confirmed for both disyllables and trisyllables. While in the shortest disyllables the contrast is "pure" (i.e., not significantly reinforced by the durations of other segments), in all trisyllables it is enhanced through the durationally inverse (compensatory) effects in other segments. In this, the situation in trisyllables is closer to that attested in other languages with ternary consonantal quantity than the situation in disyllables. The phonological quantity contrast has been lost from the second syllable vowel of trisyllables, and its duration is now inversely related to the first syllable complexity. In the segments preceding this vowel, all compensatory effects are purely phonetic. Shorter segmental durations and stronger compensatory effects in trisyllables than in disyllables indicate tendencies for both polysegmental and polysyllabic shortening. We discuss a potential relation of observed compensatory effects of shortening and lengthening (a "half-long" vowel) to foot isochrony and metrical stress.

1. Introduction

Soikkola Ingrian is a vanishing Finnic variety near St. Petersburg, Russia, with only few elderly speakers left. It is one of the four dialects of the Ingrian language, along with Lower Luga and extinct Hevaha and Oredež (Kuznetsova et al., 2015). Soikkola Ingrian features one of the most complex Finnic quantity systems which includes a rare ternary quantity contrast of consonants: *kana* [ˈka**n**aː][1] 'hen' — *kaňnā* [ˈka**n**ˑaː] 'hen:PRT' — *linnā* [ˈli**n**ːaː] 'city:PRT'. Such a contrast is attested, apart from Ingrian, only in cognate Finnic languages (Estonian and Livonian) and in slightly more distant Saami languages.

The Soikkola Ingrian ternary quantity contrast shows exceptional synchronic and diachronic features as compared to those found in cognate languages. Diachronically, the middle quantity class of consonants — short geminates — has developed in Soikkola Ingrian through the so-called secondary gemination of singletons, rather than through the shortening of primary (original) long geminates, as in Standard Estonian and Livonian, cf. the following sound changes (short geminates are in bold):

gloss	Soikkola Ingrian	Standard Estonian
'hen:PRT'	**kanā* > ***kaňnā*** [ˈka**n**ˑaː]	**kanā* > *kana* [ˈkanaː]
'city:GEN'	**linnan* > *linnan* [ˈlinːan]	**linnan* > ***liňna*** [ˈli**n**ˑa(ˑ)]

Abbreviations: 1, 2, 3, 1st, 2nd, 3rd, person; ACT, active; ALL, allative; CMP, comparative; EMPH, emphatic; ESS, essive; GEN, genitive; IMP, imperative; ILL, illative; IN, inessive; INF, infinitive; IMPS, impersonal; PL, plural; PRS, present; PRT, partitive; POS, possessive; PST, past; PTC, participle; SG, singular; SUP, supine.

[1] Italicised forms render conventional phonological transcription (quantity contrasts of vowels and consonants are marked as *a* –ā and *t* - *ťt*– *tt*), while forms in [] present phonetic transcription in IPA. In Tables 1–2 and Appendix 1, phonological transcriptions are not italicised.

https://doi.org/10.1016/j.wocn.2023.101246
0095-4470/

Synchronically, the Soikkola Ingrian ternary contrast was found not to be significantly enhanced by secondary durational cues in other foot segments, unlike in Estonian, Livonian, and Saami (Markus et al., 2013). For this reason, it has been reported as the "purest" known example of consonantal ternary quantity (Kuznetsova, 2015). This observation, however, has been made only on the basis of the disyllabic (binary) foot, but the ternary quantity contrast exists in Soikkola Ingrian also in the trisyllabic (ternary) foot (Sovijärvi, 1944). Here we adopt a definition of the foot as the metrical stress domain, delimited by rhythmically or lexically stressed syllables, whether primary or secondary (cf. Turk & Shattuck-Hufnagel, 2020, p. 29; see more in Section 2).

The Soikkola Ingrian trisyllabic foot manifests prosodic differences from the disyllabic foot. Most notably, it shows two idiosyncratic phenomena, one synchronic and one diachronic, considered to be prosodically interrelated already by an early scholar of Soikkola Ingrian, Antti Sovijärvi (1944, pp. 182–184). First, Sovijärvi impressionistically attested ongoing shortening of long vowels in the unstressecn trisyllabic foot: **pīrāta* > *pīr**a**ta* [ˈpiːr**a**d̥a] 'pie:PRT', but not in the disyllabic foot: *hār**ā*** [ˈhaːr**aː**] 'branch:PRT'. Second, diachronically, the trisyllabic foot shows unique types of secondary gemination before a short vowel, as compared to the Soikkola Ingrian disyllabic foot and to any foot type in other Finnic languages, where gemination occurs only before long vowels (see Section 6.2).

There are acoustic studies on segmental durations in the Soikkola Ingrian disyllabic foot (Kuznecova, 2009a; Markus, 2010, 2011), but the trisyllabic foot and its ternary quantity contrast have never been the object of a thorough experimental investigation so far. Our study is dedicated to an interaction between segmental length (quantity) of segments and the foot nucleus structure in 22 types of the trisyllabic foot, with a focus on consonantal ternary quantity and second syllable vowel shortening (STUDY 1). The four shortest trisyllabic types are additionally compared to four disyllabic feet with the same foot nucleus structure. Foot nucleus is a sequence from the first syllable vowel throughout the second syllable vowel, where the main length contrasts and quantity alternations are concentrated (by analogy to the syllabic nucleus). For example, the trisyllabic foot ***kattil**a* [ˈk**atːi**ɫa] 'cauldron' has the same foot nucleus structure (marked in bold; schematically VC:V[2]) as the disyllabic foot ***natta*** [ˈn**atːa**] 'slime'.

Duration of segments in Finnic and Saami languages is known to be extremely variable as a function of the number and quantity of subdomains within prosodic domains of different hierarchic levels: the syllable, the foot, the word (up to eight durational allophones are distinguished in traditional Finno-Ugric transcription; cf. Iivonen et al., 1990, p. 58). For this reason, quantity contrasts are best observed in the otherwise maximally comparable structural and prosodic contexts (Lehiste, 1968; Nahkola, 1987, pp. 240–242), and this consideration has been taken into account in the design of our study.

In Finnic languages, durational ratios have been shown to be more important for perception, as well as more stable across structural contexts in speech production than absolute durations (see overviews in Lippus, 2011; Kuznetsova, 2018a). Soikkola Ingrian is not an exception to this (Kuznecova, 2009a, pp. 64–67; Markus, 2010, 2011), and we report in our study the most important proportional ratios between the quantity classes (Section 5.1).

Particular structural effects on duration and length known for Finnic and Saami languages can be divided into two types: "compensatory" and "anticompensatory". Compensatory effects imply an inverse relation between the duration of a given prosodic subconstituent (e.g., segment, syllable, foot) and the length and number of subconstituents within a given constituent (syllable, foot, word, respectively). Anticompensatory effects imply a direct relationship of the same kind. This paper considers the compensatory effects (see specific hypotheses in Section 3), while the anticompensatory effects are briefly mentioned in Section 5.3.3. Phonological secondary geminates historically emerged out of phonetic anticompensatory lengthening (Kuznetsova, accepted). However, as we show in this paper, synchronic secondary geminates undergo the same compensatory effects as other quantity classes.

The compensatory effects primarily include durational shortening at different prosodic levels, when durations decrease by adding more segments, syllables, or feet to a respective prosodic domain. This shortening has recently received an umbrella term "poly-subconstituent shortening" (cf. Turk & Shattuck-Hufnagel, 2020, pp. 135–143 for a general overview). For Soikkola Ingrian, compensatory shortening at different prosodic levels was briefly mentioned already by Sovijärvi (1944, pp. 11–12). In our study, both polysegmental (Section 5.4) and polysyllabic (Section 5.3) shortening is discovered, and both are found stronger in the trisyllabic foot than expected from previous studies on the disyllabic foot. Most importantly, we find that the ternary quantity contrast of consonants in the trisyllabic feet is robust (Section 5.1) but, unlike in the (shortest) disyllabic feet, it is significantly enhanced by the durations of other foot nucleus elements (Section 5.3.2).

Known compensatory effects in Finnic languages also include a phenomenon which can be interpreted as compensatory lengthening. In the foot with the stressed light syllable (C)V (i.e., an open syllable with a short vowel), referred to below as "light foot", the unstressed second syllable short vowel undergoes phonetic lengthening: *kana* [ˈkanaː] 'hen'. This effect has been called "half-long vowel" since Hakulinen (1922), and we confirm it here also for the Soikkola Ingrian trisyllabic foot (Section 5.2).

Contrarily, in the "heavy foot" with the stressed heavy syllable (of any other structure than (C)V, i.e., containing a long vowel or/and closed; for example, *kaňnā* [ˈkanˑaː] 'hen:PRT'), both long and short second syllable vowels frequently undergo shortening or qualitative reduction in Finnic varieties (cf. overviews in Kuznetsova, 2016; Kuznetsova & Verkhodanova, 2019). A particular finding of our study is that Soikkola Ingrian manifests phonological shortening of second syllable long vowels in all types of the trisyllabic heavy foot except the shortest one (Section 5.2). This sound change can be classified as the phonological outcome of phonetic polysyllabic shortening. It has led to the complete loss of the vocalic quantity contrast

[2] Hereafter, C means any consonant, R — a sonorant, and V — a vowel. The three degrees of consonantal length are marked as C — Cˑ — C:, while the binary length contrast of vowels as V — V: (the marking V: means either a long vowel or a diphthong in the case of V_1, if not specified otherwise, but in Appendix 1 and Tables 2 and 4a, diphthongs are specifically marked as V_1V_2). In formal phonological accounts positing a two-layered trisyllabic foot (e.g. Martínez-Paricio & Kager, 2015, 2021), the "minimal foot" roughly corresponds to our disyllabic foot nucleus and the "maximal foot" to our trisyllabic foot.

in the second syllable of trisyllables, while disyllables still retain this contrast (Section 5.3.1).

The structure of the paper is as follows. In Section 2, we summarise relevant information on segmental phonology, quantity patterns, and stress in Soikkola Ingrian. In Section 3, we outline our specific hypotheses regarding the expected compensatory effects. Section 4 contains information on data and methods, and Section 5 explicates the results. Discussion in Section 6 summarises the findings and elaborates on related theoretically important issues (ternary quantity, the “half-long” vowel, stress, foot isochrony). The concluding Section 7 provides a brief outlook on the potential relevance of our results for formal and articulatory phonology.

2. Basic facts on Soikkola Ingrian quantity and stress

Soikkola Ingrian is divided into two subdialects: north and south. The north subdialect, studied here, is phonologically slightly more archaic than the south one: it exhibits fewer effects of phonological V_2 shortening, of mid vowel raising, and of singleton consonant voicing (Kuznecova, 2009a). Soikkola Ingrian has palatal vowel harmony of the Finnish type. Its phonemic system is given in Table 1 (revised from Kuznecova, 2009a). The ternary quantity contrast of consonants, which comprises singletons C, short geminates Cˑ, and long geminates Cː, exists in the intervocalic position for all consonants (cf. a triplet for *n* in Section 1). After a sonorant *l*, *m*, *n*, *r* before a vowel, this contrast is also attested for consonantal types *k*, *p*, *t*, *t′*, *s*, *s′*, e.g., *kartās* [ˈkard̥aːʒ̊] ‘map:IN’ — *kartt̆ā* [ˈkar**t**ˑaː] ‘baking_sheet:PRT’— *karttā* [ˈkar**t**ːaː] ‘map:PRT’. Singletons of these consonants and of laryngeal *h* are phonetically (half–) voiced, e.g., *mātā* [ˈmaːd̥aː] ‘sleep:IMPS’ — *lōtt̆ā* [ˈɫoː**t**ˑaː] ‘broom:PRT’— *nōttā* [ˈnoː**t**ːaː] ‘seine:PRT’, but only fricatives *v : f* manifest an emerging true voicing contrast. In this paper, the ternary quantity contrast is studied only for stops *k*, *p*, *t*.

Table 2 presents a summary of the phonological quantity patterns addressed in this study. Appendix 1 provides a broader and more detailed outlook at the main quantity patterns attested in the Soikkola Ingrian monosyllabic, disyllabic, and trisyllabic feet, with glosses and notes to particular cases. These patterns, attested as synchronic in sources covering a period from the 2nd half of the 19th century to the 1st half of the 20th century (Porkka, 1885; Sovijärvi, 1944; Nirvi, 1971), have been taken as a starting point for our study.

Due to the different positions of secondary gemination and other historical factors (see Kuznetsova, accepted), the combinations of length in C_2 and V_2 do not coincide in disyllabic and trisyllabic feet. Attested combinations of the phonological length classes in the second syllable vowel (V_2) and the consonant at the 1st/2nd syllable boundary (C_2) after the original light (C)V-, heavy (C)Vː- / (C)VR-, and extra-heavy (C)VːR- syllables are summarised in Table 3. Phonological changes revealed in these original patterns by our study are shown in Table 4 in Section 6.1.

Stress in Soikkola Ingrian has been scarcely studied (Porkka, 1885, pp. 36–37; Laanest, 1993, p. 57; Gordon, 2009). Early authors described it as a general Finnic system (first proposed for Finnish by Genetz, 1881) with primary stress on the word-initial syllable and secondary stresses on subsequent odd syllables apart from the final one. However, the general structure of metrical prominence in Finnic languages is still under-researched. Based on our earlier observations (Kuznecova, 2009b, 2009a), we describe the stress system of Soikkola Ingrian as similar to that of Estonian (viz. Kuznetsova, 2018a). The lowest type of metrical prominence is foot-initial stress. Foot stress is primarily based on specific durational cues within the foot nucleus: lack of reduction and shortening in stressed vowels and the compensatory effects mentioned in Section 1 for the unstressed vowels of the second syllable: lengthening in the light foot vs. shortening and reduction (of *a*, *ä*, *e* to [ə]) in the heavy foot. Vowels under foot stress also serve as anchors for high pitch. In a careful realisation, pitch can mark each foot (see high pitches on both foot-initial syllables in ˈ*starikale* [ˈʃtariːˌg̊aɫeː] ‘old_man:ALL’ in Kuznecova, 2009b, p. 32). More typically, however, pitch marks prominence at higher prosodic levels, such as word or accentual phrase. For example, Gordon (2009) described higher pitch and intensity, as well as longer duration of long vowels, as cues for syllables allegedly carrying primary lexical stress in Soikkola Ingrian, as compared to those under secondary stress (i.e. foot stress in absence of word stress) or without stress.

In Finnic languages, most pitch manipulations can also actually be the postlexical phenomena marking prominence at the accentual phrase and higher levels, as suggested by Gordon (2009, pp. 98–99) for Soikkola Ingrian on the basis of Suomi et al. (2003) and by Kuznetsova, 2018a for Estonian. In this sense, foot stress is a more robust phenomenon than word-level stress (as also stated for Estonian in Asu-Garcia et al., 2016, p. 160). A discussion on a possible relation between the compensatory effects observed in our study, foot stress, and foot isochrony is developed in Sections 6.3–6.4.

Foot stress can be rhythmic or lexicalised. Lexicalised stress is defined as the phonological accent (Kuznecova, 2009b; van der Hulst, 2010, 2014) and marked in phonological transcription as /´/. The phonetic correlates of rhythmic and lexicalised stress are the same. The main difference is that lexicalised stresses are bound to certain syllables in certain morphemes (roots, some inflectional and derivational affixes) and cannot shift to another syllable, unlike rhythmic stresses. All stretches of speech longer than two syllables between lexicalised accents are additionally rhythmically divided into feet, mostly trochaic (disyllabic) and dactylic (trisyllabic). Monosyllabic feet with a truncated foot nucleus exist either as monosyllabic words (cf. Appendix 1) or as word-final monosyllabic feet. The placement of rhythmic stresses is to a certain degree variable in multifoot words (viz. Kuznecova, 2009b). For example, in *´naise´likkaha(l̆)le* ‘married:ALL’, the first two stresses are the lexicalised accents bound to the root *´naise-* and the suffix *´- likka-* and the last one is rhythmic stress with variable placement: [ˈnai̯ʒ̊ĕˌlikːăˌhaɫˑe] ∼ [ˈnai̯ʒ̊ĕˌlikːăɣăˌɫeˑ] (see also Section 6.3).

3. Research questions of this study

Our specific hypotheses regarding the expected types of compensatory effects in segmental durations are primarily based on ealier studies of Soikkola Ingrian, but where relevant Soikkola Ingrian data were not available we drew upon studies on other Finnic languages. We look at the three key elements

Table 1
Soikkola Ingrian phonemes.

(a) Vowels			
Height	Backness		
	Front		Non-front
	Unrounded	Rounded	
High	i ī	ü ǖ	(ɨ)[a] u ū
Mid	e ē/ī [b]	ö ȫ/ǖ	o ō/ū
Low	ä ǟ		a ā

(b) Consonants					
Manner		Place			
		Labial	Dental	Palatal(ised)	Velar
Stops		p p̆p pp	t t̆t tt	(tʲ t̆tʲ ttʲ)	k k̆k kk
Fricatives		(f f̆f ff)[a]	s s̆s ss	(sʲ s̆sʲ ssʲ)	h h̆h hh
Sonorants	Nasals	m m̆m mm	n n̆n nn	(nʲ n̆nʲ)	
	Laterals		l l̆l ll	(lʲ l̆lʲ)	
	Trills		r r̆r rr	(rʲ r̆rʲ)	
Affricates				ʧʲ ttʃʲ	
Glides		v v̆v vv		j j̆j jj	

Notes: [a]Phonemes in parentheses occur only in expressive (i.e., onomatopoetic, diminutive) lexemes and/or in Russian loans. [b]Long mid vowels in the stressed root syllables are realised in a range from mid to high depending on a village and a speaker.

Table 2
Original quantity patterns in the nuclei of 22 studied disyllabic and trisyllabic structures[a].

Syllable number	Foot nucleus	Singleton $C_2 + V_2$	Short geminate $C_2 + V_2$		Long geminate $C_2 + V_2$	
		C + V	C· + V	C· + V:	C: + V	C: + V:
2 syllables	VCV	ˈtapa		ˈtap̆pā	ˈnatta	ˈtappā
3 syllables	VCV	ˈlakata		ˈmat̆tāla	ˈkattila	ˈkattīma
	V:CV	ˈvīkate	ˈvūt̆tava	ˈsūt̆tīma	ˈūttele	ˈmūttīma
	V_1V_2CV	ˈleikata	ˈvoit̆teli	ˈhoit̆tīma	ˈvoitteli	ˈtoittīma
	VRCV	ˈharkata	ˈmurk̆kina	ˈkerk̆kīmä	ˈmarkkoja	ˈharkkāma
	V:RCV	ˈkērsimä[b]	ˈvǟnt̆teli	ˈvǟnt̆tīmä		

Notes: [a]"CV" refers to one of the $C_2 + V_2$ length class combinations in the next three columns; grey-shaded cells represent structurally impossible combinations. [b]In our study, a structure with a long third syllable vowel (*väntelö*) was used instead, cf. Appendix 1 and Section 4.

Table 3
Original combinations of C_2 length and V_2 length after (C)V-, (C)V:- / (C)VR-, and (C)V:R- in the first syllable.

Foot type	$C_2 + V_2$	C_2 + V:$_2$	C·$_2 + V_2$	C·$_2$ + V:$_2$	C:$_2 + V_2$	C:$_2$ + V:$_2$
Disyllabic	✓	✓ but not after (C)V-	×	✓	✓ but not after (C)V:R-	
Trisyllabic	✓	×	✓ but not after (C)V-	✓		

of the foot nucleus: the first syllable vowel or diphthong (V_1),[3] the second syllable vowel (V_2), and the intervocalic consonant (C_2) or a consonant cluster. As noted in Section 2, in Soikkola Ingrian, the ternary quantity contrast of stops exists in C_2 in only one type of clusters: after a sonorant (R), so this element was also included in the study. Positions outside the foot nucleus (word-initial consonant C_1 and the segments of the third syllable) do not fall within the scope of this paper (cf. Kuznetsova, accepted; Kuznetsova et al., 2023 for some data on these positions).

Vocalic and consonantal classes of greater length show more durational variability, including stronger compensatory effects, than the shorter ones in Finnic languages (cf. Sadeniemi, 1949, p. 104; Lehtonen, 1970, p. 35; Nahkola, 1987, pp. 15–16 for Finnish). This is in line with an observation that variability in both speech and non-speech movements grows linearly with an increase in duration of an interval (viz. Turk & Shattuck-Hufnagel, 2020, pp. 90–95). Expectations about length contrasts in the first syllable of the Soikkola Ingrian trisyllables are formulated as our first research question (RQ1). In Finnic languages, the duration and length of the second syllable vowel (V_2) generally manifests stronger compensatory effects than the duration and length of preceding elements. Structural influences on V_2 duration and length in trisyllables are addressed as RQ2. Polysyllabic shortening in the presence of the third syllable in trisyllables, as compared to disyllables, is investigated for the four shortest foot nucleus structures as RQ3. Polysegmental shortening effects of higher

[3] Finnic diphthongs in V_1 are comparable in their duration and overall prosodic properties to long monophthongs (Lehtonen, 1970, p. 69; Eek & Meister, 2004), which was confirmed in our study regarding the compensatory effects. Therefore, long monophthongs and diphthongs are generally considered together (as long V_1) in this paper.

Table 4
Synchronic phonological outcome of STUDY 1–2, as compared to the original situation reflected in Tables 2-3.

(a) New quantity patterns in the nuclei of studied disyllabic and trisyllabic structures (durationally long V_2 are in bold)

Syllable number	Foot nucleus	Singleton $C_2 + V_2$	Short geminate $C_2 + V_2$		Long geminate $C_2 + V_2$	
		C + V	Cˑ + *V	Cˑ + *Vː	Cː + *V	Cː + *Vː
2 syllables	V<u>CV</u>	tap**a** [ˈtab̥**a**ː]		tap̆p**ā**	natta	≠ tapp**ā**
3 syllables	V<u>CV</u>	lak**a**ta [ˈłag̊**a**ːd̥a]		mat̆t**ā**la	kattila = *kattīma	
	Vː<u>CV</u>	vīkate	vūt̆tava = *sūt̆tīma		ūttele = *mūttīma	
	V_1V_2<u>CV</u>	leikata	voit̆teli = *hoit̆tīma		voitteli = *toittīma	
	VR<u>CV</u>	harkata	murk̆kina = *kerk̆kīmä		markkoja = *harkkāma	
	VːR<u>CV</u>	kērsimä	vānt̆teli = *vānt̆tīmä			

(b) New combinations of the C_2 length and V_2 length after (C)V-, (C)Vː- / (C)VR-, and (C)VːR- in the first syllable

Foot type	$C_2 + V_2$	$C_2 + Vː_2$	$Cˑ_2 + V_2$	$Cˑ_2 + Vː_2$	$Cː_2 + V_2$	$Cː_2 + Vː_2$
Disyllabic	✓	✓ but not after (C)V-	×	✓	✓ but not after (C)VːR-	
Trisyllabic	✓	×	✓ but not after (C)V-	✓ only after (C)V-	✓ but not after (C)VːR-	

quantity or extra number of adjacent segments on the duration of the foot nulceus elements preceding V_2 in trisyllables are discussed as RQ4. Specific hypotheses for each RQ are outlined below.

3.1. RQ1: Length contrasts in V_1 and C_2 of trisyllables

A binary quantity contrast in V_1 and a ternary quantity contrast in C_2 have been reported for the Soikkola Ingrian disyllabic foot (Sovijärvi, 1944, p. 12; Kuznecova, 2009a; Markus, 2010, 2011; Markus et al., 2013). On the basis of limited earlier measurements of the shortest trisyllables and impressionistic observations (Kuznecova, 2009a, p. 64), we suggested that, in otherwise maximally comparable structures, binary length in V_1 will be robust in all structural types of trisyllables, while the ternary length in C_2 will be robust at least in the shortest trisyllables. The situation with ternary length in longer trisyllabic structures was not entirely clear. These expectations are formulated as hypotheses **H1A** and **H1B**:

- **H1A**: in V_1 of trisyllables, a robust contrast of binary length is always expected in comparable structures.
- **H1B**: in C_2 of trisyllables, a robust contrast of ternary length is expected at least in the shortest structures.

3.2. RQ2: Compensatory effects of foot structure in V_2 of trisyllables

Expected compensatory effects in V_2 differ between the light and the heavy foot. The light foot in Soikkola Ingrian can only have the VCV nucleus, because all former light feet with the *VCVː nucleus became heavy through secondary gemination (**kanā* > *kaňnā*). Short V_2 in the VCV nucleus undergoes phonetic compensatory lengthening (first discussed for this dialect in Sovijärvi, 1944, p. 10). In prosodically innovative Finnic varieties, to which Soikkola Ingrian belongs, this lengthened short V_2 ("half-long vowel", cf. Section 6.3) is very prominent and usually has longer duration than phonologically long V_2 in the heavy foot (Kuznetsova, 2016).

In the heavy foot, all V_2 in these innovative Finnic varieties undergo more or less strong compensatory shortening. Soikkola Ingrian is still relatively conservative in this respect as compared to adjacent Finnic varieties: phonologically long V_2 are preserved in disyllables (Markus, 2011) and any qualitative reduction of short V_2 to schwa is still occasional and purely phonetic (Kuznecova, 2009b). However, for the trisyllabic heavy foot, Sovijärvi (1944, pp. 182–184) indicated ongoing phonological shortening of long V_2, cf. also Kuznecova (2009a). Such shortening was not yet attested in the second half of the 19th century (Porkka, 1885; cf. "*kiukaada*" 'oven: PRT' without long V_2 reduction on p. 68) and might be a relatively recent development. However, its entire scope in terms of degree and the structures affected in modern Soikkola Ingrian was not clear, which was one of major motivations for the present study. The following expectations about compensatory effects in V_2 are tested:

- **H2A**: in any light foot, stress-induced lengthening of phonologically short V_2 is expected.
- **H2B**: in the heavy trisyllabic foot, phonological shortening of long V_2 is expected at least in some structures.

3.3. RQ3: Polysyllabic shortening in the foot nucleus of trisyllables as compared to disyllables

Earlier studies suggested shorter durations of V_1, V_2, and long geminates C_2 in trisyllables as compared to disyllables for Estonian, Livonian, and Inari Saami (Lehiste, 1968, 1997; Lehtonen, 1970; Tuisk, 2015; Türk et al., 2020). The shortening of long V_2 attested by Sovijärvi in trisyllables (cf. Section 3.2) suggests the existence of polysyllabic shortening also for Soikkola Ingrian, because its disyllabic structures did not show any trend of this kind (Kuznecova, 2009b; Markus, 2011). Shorter durations of primary and secondary geminates in trisyllabic than in disyllabic feet were reported for this variety by Sovijärvi (1944, pp. 11–12) and Gordon (2009, p. 93), and of "half-long" V_2 by Sovijärvi (1944, p. 17). Our general expectation about the polysyllabic shortening in trisyllables as compared to disyllables is coded as H3. Its sub-hypothesis 1 is specifically related to the phonological difference in V_2 quantity between trisyllables and disyllables as an outcome of this polysyllabic shortening:

- **H3**: segmental durations are expected to be shorter in the foot nuclei of trisyllabic feet, as compared to those in the identical foot nuclei of disyllabic feet.
  - **Sub-hypothesis 1:** long V_2 manifests phonological shortening in trisyllables but not in disyllables.

In earlier studies (Markus, 2011), the shortest Soikkola Ingrian disyllables, e.g. in a triplet like *tapa* [ˈtab̥aː] ‘kill:IMP’ — *tap̆pā* [ˈtap·aː] ‘catch:3SG’ — *tappā* [ˈtapːaː] ‘kill:3SG’, manifested the lack of statistically significant inverse durational ratios of the second syllable vowels to consonantal quantity (unlike in Estonian, Livonian, and Saami). However, because of the ongoing phonological shortening of long V_2, the situation might be different in this respect for trisyllables, although we do not know exactly how. Our study aimed at finding this out as well. Our general expectation is formulated below as the second sub-hypothesis of H3:

- **Sub-hypothesis 2:** an inverse relation of V_2 duration to C_2 length in the shortest disyllables is expected at most as a weak trend, but for comparable trisyllables the situation might be different.

3.4. RQ4: Polysegmental shortening in the elements preceding V_2 in trisyllables

In Soikkola Ingrian disyllables, long vowels V_1 and the following long and short geminates C_2 shortened significantly each other’s durations (Markus, 2011, pp. 112, 114; see also Gordon, 2009). Long vowels were shorter before both types of geminates than before singletons, while geminates were longer after short vowels than after long vowels. This is typical of Finnic languages (e.g. Eek, 1975; Lehtonen, 1970; Lippus et al., 2013; Suomi et al., 2013) and was expected also for Soikkola Ingrian trisyllables.

Structures with a sonorant R before C_2 have not been studied for Soikkola Ingrian. However, limited data on Standard Estonian also showed compensatory effects. The presence of a sonorant R between long V_1 and geminate C_2 shortened the duration of both (Eek, 1975, pp. 50, 52), while the gemination of C_2 led to the shortening of the sonorant itself (Eek, 1975, p. 52). Eek’s measurements (1975; Eek & Meister, 2004) often indicated at least a weak trend for compensatory shortening of R in the proximity of long V_1 and geminate C_2, as compared to similar structures with short V_1 and singleton C_2.

The duration of short V_1 and singleton C_2 either did not significantly change under structural influences in all aforementioned studies, including those on Soikkola Ingrian, or showed a weak trend towards the same kind of compensatory shortening as outlined above for their respective long counterparts.

The following hypotheses and sub-hypotheses are formulated for polysegmental shortening in the foot nucleus elements preceding V_2:

- **H4A**: long phonemes (long V_1 and long and short geminate C_2) are expected to manifest significant phonetic shortening in the context of increased quantity and number of other first syllable elements.
  - **Sub-hypothesis 1:** shortening of long V_1 and geminate C_2 is expected in the context of increased quantity of C_2 and V_1, respectively, as compared to the contexts with their shorter length classes.
  - **Sub-hypothesis 2:** shortening of long V_1 and geminate C_2 is expected in the presence of a sonorant R in the foot nucleus, as compared to its absence.
- **H4B**: short phonemes (short V_1, singleton C_2, and short sonorant R) are not expected to change their duration significantly as a function of the first syllable structure.
  - **Sub-hypothesis 1:** short phonemes might exhibit a weak compensatory trend in the same conditions as outlined for long phonemes in H4A.

The final note regards secondary geminates. As mentioned in the introduction, they have originally emerged in Soikkola Ingrian through an anticompensatory effect, which is opposite to the compensatory phenomena addressed in this paper. However, as shown in several studies, phonologised secondary geminates are similar in prosodic properties to primary geminates. In particular, they affect the neighbouring sounds and are affected by them as long phonemes and in a compensatory manner. For example, the longer the duration obtained in Finnish dialects by phonologised secondary geminates, the stronger the shortening of the following and preceding vowels (Sovijärvi, 1944, p. 27; Lehtonen, 1970, pp. 126–129; Palander, 1987, pp. 222–229; Nahkola, 1987, pp. 257–260). Conversely, secondary geminates in structurally more complex feet are themselves phonetically shorter than in less complex feet in Soikkola Ingrian (Sovijärvi, 1944; Gordon, 2009).

Therefore, we do not expect Soikkola Ingrian secondary (short) geminates to show any extraordinary features as compared to primary (long) geminates in regard to hypothesis **H4A**.

4. Data and methods

STUDY 1 included two questionnaires with 22 structural types of trisyllables exemplified in Table 2 (altogether 3812 tokens; see details in Appendix 2). Out of the 22 types, one was a bifoot structure *vǟntelȫ* ‘twist:3SG’ with a final long vowel, to cover the foot nucleus VːRCV (see Ftn. (c) to Appendix 1). A consonant in the beginning of the third syllable (C_3) was a singleton where available, otherwise a short geminate (see Appendix 2). Each structure in Questionnaires 1–2 was included in five phrases with different carrier words, and each phrase was pronounced from 4 to 7 times by each of the five speakers. STUDY 2 (Questionnaire 3) compared four disyllabic feet exemplified in Table 2 (429 tokens) with respective trisyllables of the same foot nucleus structure (first row of trisyllables in Table 2; 383 tokens) in three of the five speakers. Each of the four disyllabic structures included 7–12 tokens pronounced by speakers from two to twelve times. The full set of token counts for each wordform in each speaker in STUDY 1–2 is given as Supplementary material. All collected tokens were in phrase-final position.

Data were collected in the Kingisepp district (Russia) in 2014–2016 in the residences of the five speakers — all female, who learned a north (more conservative) subdialect of Soikkola Ingrian (or a transitory north-to-south variety of *Vīstinä*) in their childhood. Older speakers ST, LM and EI still spoke Ingrian with their family members and friends at the time of the interview and considered it as their only mother tongue. Younger speakers AL and AG were less fluent and considered both Ingrian and Russian as their mother tongues. Speakers were presented with a phrase spoken in Russian, asked to translate it into Ingrian, asked to change it if a carrier word had not been pronounced in sentence-final position (word order is free in Ingrian), and then to repeat it three or four times in moderate tempo.

Data were semi-automatically segmented in Praat (Boersma & Weenink, 2013) and manually corrected. Raw durations were extracted with the help of Python and Praat scripts and statistically analysed in R (v. 4.0.0, R Core Team, 2020) with the use of *lme4* (Bates et al., 2015), *car*, *effects* (Fox & Weisberg, 2019), *emmeans* (Length, 2020), *ggplot2* (Wickham, 2016), and *lmerTest* (A. Kuznetsova et al., 2017) packages. Mixed effects linear regression models were fitted to study relationships between the structural parameters of the foot nucleus and the durations of foot nucleus elements. Main fixed effects were the length classes of all segments in the foot nucleus and the presence of a sonorant in the first syllable coda. Originally a distinction between long monophthongs and diphthongs in V_1 was also included where possible, but it did not emerge as a significant predictor for any compensatory effect and is not discussed here.

These effects were checked, where relevant, in a binary interaction with segmental length in each studied position. A ternary interaction in the first syllable was studied only for V_2, where a prediction was available (cf. Section 5.2). Random effects included by-speaker and by-word random intercepts (Baayen et al., 2008), as well as phonemic qualities in a given position (cf. Lehiste, 1960; Liiv, 1961; Eek, 1974; Lehtonen, 1970, pp. 60–75; Suomi et al., 2008 on intrinsic segmental durations) and in the two adjacent positions (Liiv, 1961; Lehtonen, 1970, pp. 76–81). For the models on V_1, the presence vs. absence of a preceding word-initial consonant C_1, which trivially shortens vs. lengthens V_1 duration, was also included. A random slope was included in only one case where a prediction was availaible. In case of model convergence failure, insignificant random effects concerning phonemic qualities were excluded from some of the maximal models following backwards stepwise elimination procedure in *lmerTest*. Phonemic quality effects were also not included in case of high collinearity with fixed effects. Final models are reported in the respective subsections of Section 5.[4]

Effects usually returned significant *p*-values (<0.05) at *t*-values $\geq |2|$. However, a significant *p*-value does not necessarily mean that an effect is phonological. Robust length contrasts usually had a *t*-value >|5| across our tests, and smaller significant values generally indicated phonetic compensatory or anticompensatory effects. Results were also placed against the minimal perceptibility threshold for phonetic durational differences (based on Weber's law, a non-linear measure which depends on the absolute duration of compared units). For English foot durations between 300–500 ms, this threshold has been assessed as 30–100 ms in optimal (experimental) hearing conditions and even higher in the normal ones (Lehiste, 1977, p. 258). Modern neuroimaging also shows that a brain response to the categorical phonemic contrasts generally starts at about 50–100 ms (e.g. Obleser & Eisner, 2009). Phonemic length contrasts, therefore, are supposed to have at least a minimum of 30–40 ms of durational difference.

Durational variability of sounds is also known to be affected by speech tempo and by phrasal effects. However, Estonian and Finnish data showed that variable speech tempo does not distort proportional durational relations between sounds (Eek & Meister, 2003; Suomi et al., 2013). Moreover, we studied the dependency of duration on the length and/or presence of segments rather than dependencies between durations, where speech tempo would be indeed a crucial factor. The phrase-final position was chosen because length contrasts are articulated phrase-finally very clearly, which allowed us to avoid a risk of occasional vowel contrast reduction or merger caused just by an unaccented phrasal position. However, this position is known to cause certain prosodic effects on segmental durations, which is discussed in Section 6.4.

5. *Results of* STUDY 1–2

Results pertinent to each of the hypotheses in Section 3 are organised according to the four RQs (Sections 5.1-5.4 respectively). The synthesis of the main findings is given in Section 6.1.

5.1. RQ1: Length contrasts in the V_1 and C_2 of trisyllables (STUDY 1)

Both V_1 and C_2 manifest robust length contrasts: the binary length contrast of vowels (modelled[5] 94 vs. 181 ms***, cf. Appendix 6) and the ternary length contrast of consonants (modelled 94 vs. 177 vs. 234 ms***), with ratios 1 : 2 and 1 : 2 : 2.5, respectively. Aggregated raw and modelled mean durations for these length contrasts are given in Fig. 1, while Appendix 4b–c presents detailed data on the raw mean durations of V_1 and C_2 in all 22 foot nucleus types. The durations of singletons and short geminate stops in C_2 are close to the durations of short and long vowels in V_1, respectively, while long geminate C_2 stops can be called phonetically overlong.

Greater quantity of both V_1 and C_2 shows progressively higher durational variability, cf. a progressively growing size of "boxes" (i.e., the ranges from the first to the third quartile) in the progressively greater length classes of V_1 and C_2 in Fig. 1a-b. This variability is at least in part linked to the progressively higher sensitivity of longer sounds to structural influences (discussed in Section 5.4).

However, all structural influences on segmental durations in the V_1 and C_2 of trisyllables are purely phonetic. In maximally comparable contexts, all the quantity classes of V_1 or C_2 are always clearly distinguished. As for C_2, we expected this for the shortest trisyllables, but were not sure about the longer ones (cf. doubts expressed by Kuznecova, 2009a, pp. 67–68). Our present data confirm the ternary contrast for all types of studied trisyllabic structures. Fig. 2 gives an illustration to our conclusion that ternary length is always distinguished well in structurally comparable contexts (i.e. differing only in the length of C_2), although it might superficially look like a "continuum of lengthening" (cf. Gordon, 2009) if the structural factors are not taken into

[4] Acoustic measurements and scripts used in this study will appear at: https://osf.io/rbw3m/.

[5] Models for V_1 and C_2 in trisyllables are given in 5.4 in relation to the phonetic compensatory effects in their durations.

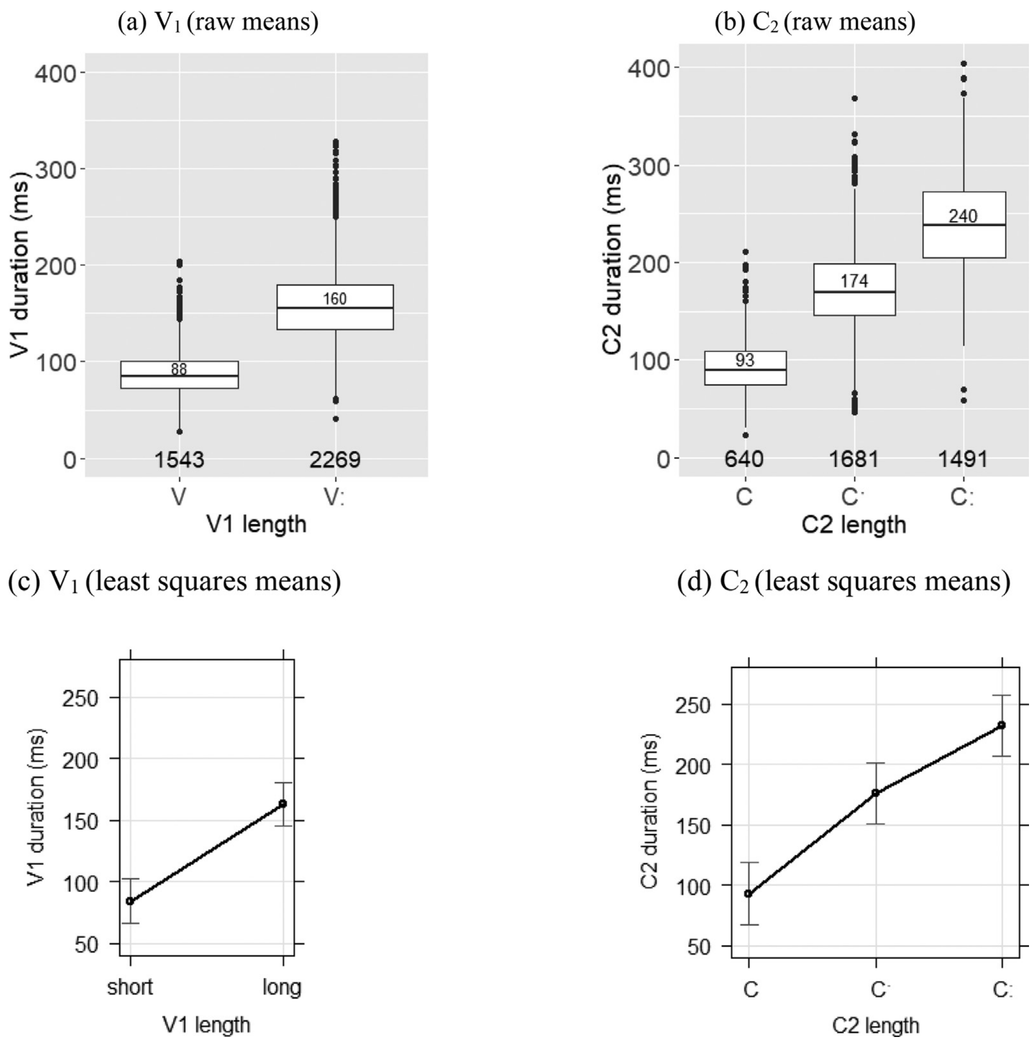


Fig. 1. Duration (in ms) of the length classes of V_1 and C_2 in trisyllables. In (a-b), raw means (in ms) are reported within boxes; the number of tokens is given below each box at the bottom.

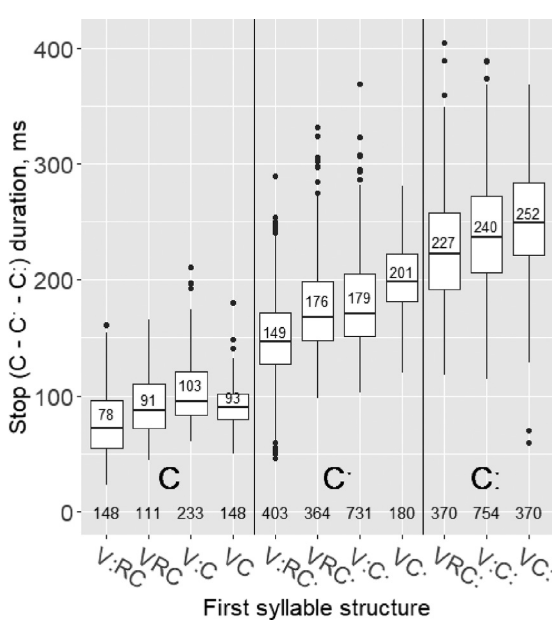


Fig. 2. Raw mean duration (in ms) of the three length classes of C_2 stops in Soikkola Ingrian grouped by the existing combinations of V_1 length, C_2 length, and the presence of a sonorant R. Raw means (in ms) are reported within boxes; the number of tokens is given below each box at the bottom.

account. In Fig. 2, we grouped together the structures by the main combinations of the structural factors in the first syllable (V_1 and C_2 length and the presence of R). It can be seen that the duration of C_2 grows in a perfectly inverse relation to the decrease in structural complexity and segmental quantity in the first syllable (with an exception of short C_2 in the light foot VCV). Compensatory impact of the foot nucleus structure to C_2 duration is addressed in detail in Section 5.4.

5.2. RQ2: Restructuring of the V_2 length contrast in trisyllables (STUDY 1)

The hypotheses under RQ2 test compensatory phonetic lengthening of short V_2 in the light foot (**H2A**) and compensatory shortening of long V_2 in the heavy foot, expected at least in some structures (**H2B**).

Phonologically long V_2 after a heavy syllable has indeed undergone significant shortening. This process has changed the original phonological contrasts of V_2 in trisyllables. Four durational types of V_2 distinguished in synchrony in the studied trisyllables are represented in Fig. 3 (see Appendix 4a for a more detailed outlook at all 22 structures):

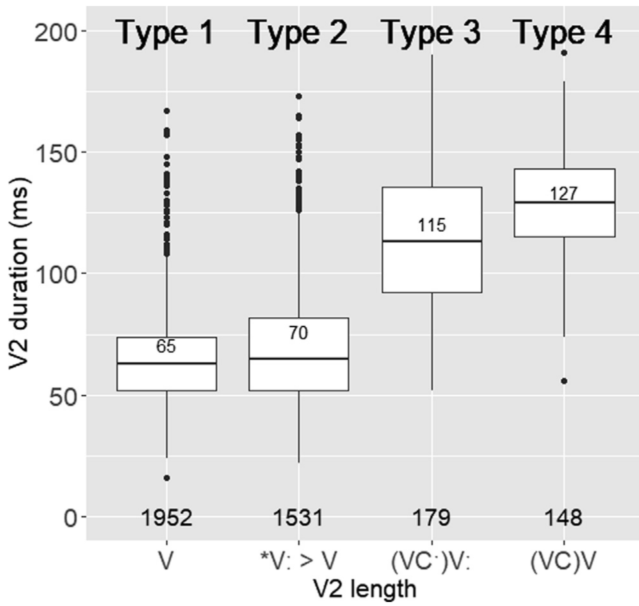


Fig. 3. Four synchronic durational classes of V_2 in trisyllables. Raw means (in ms) are reported within boxes; the number of tokens is given below each box at the bottom.

Type 1 ("V", e.g., *katt**i**la* 'cauldron'): original short V_2 after a heavy syllable — remain short;

Type 2 ("*V: > V", e.g., **kattīma* > *katt**i**ma* 'cover:1PL'): original long V_2 after a heavy syllable (apart from the VC·V: nucleus) — have nearly shortened (a support for **H2B**);

Type 3 ("(VC·)V:", e.g., *maťť**ā**la* 'low'): phonologically long V_2 in the VC·V: nucleus, the shortest type of nuclei with a phonologically long V_2 after a heavy syllable — remain long;

Type 4 ("(VC)V", e.g., *lak**a**ta* 'sweep:INF'): phonologically short V_2 in VCV, after a light syllable — are phonetically lengthened (the "half-long" vowel, cf. Section 6.3; a support for **H2A**).

Hypotheses **H2A** and **H2B** predict opposite effects on V_2 of the heavy vs. the light foot: shortening of long V_2 in the former while lengthening of short V_2 in the latter (the "half-long" vowel). This expected contrast between V_2 in the light foot and V_2 in those heavy foot types which are considered in our STUDY 1 can be accounted for with a three-way interaction of the three main predictors in the first syllable (V_1 length, C_2 length, and the presence of R; the matrix is incomplete as a combination *V:RC: does not exist), which was included in the model. For V_2, our qualitative exploration (Kuznetsova & Markus, 2022) has shown interspeaker differences in the degree of long V_2 shortening, so the original phonological length of V_2 from Table 2 is included in the model both as a random intercept and as a speaker-specific slope: V_2 duration ~ V_1 length * Presence of R * C_2 length + V_2 etymological length + (1 + V_2 etymological length | speaker) + (1 | wordform) + (1 | V_2 quality) + (1 | C_2 quality), data = 3812 tokens. The ternary interaction is significant for V_2 duration (p = 0.008**), but the original length of V_2 is not (8.5 ms, n.s.; cf. Appendix 6).

Our hypothesis **H2B** does not predict any specific differences between the durations of long V_2 after different structural types of the first heavy syllable, although a generally incomplete status of long V_2 shortening is expected. Fig. 3 shows that the duration of long V_2 after one type of the stressed heavy syllable (Type 3) stands out in comparison with long V_2 after all other stressed heavy syllable types (aggregated in Type 2; see also Appendix 4a). This outstanding type (Type 3) represents V_2 after the shortest type of the stressed heavy syllable (VC·), as in *maťť**ā**la* 'low'. The difference between this type and all the other studied heavy first syllable types is also modelled by the aforementioned three-way interaction. Specific phonological and historical features of the foot nucleus type VC·V: are addressed in Section 6.3.

Least squares mean duration of V_2 after VC vs. after VC· (i.e., Types 3 vs. 4) in our model differ by 22 ms***[6]. Additionally, the V_2 of both Type 3 and 4 show differences from V_2 durations in the foot nuclei with all other combinations of the three predictors. Modelled duration of V_2 in VCV differ from these other types by 58–66 ms*** and in VC·V:, by 35–43 ms***. Contrarily, modelled V_2 durations across other types of the first syllable structure do not show any significant mutual differences.

These results confirm the hypotheses **H2A** and **H2B**, also adding more clarity to the process of long V_2 shortening in Soikkola Ingrian. The restructuring of the V_2 length contrast in trisyllables is still in progress, observed as such also by Sovijärvi (1944, pp. 182–184), but in fact this contrast can be considered lost, for two reasons summarised in Section 6.1.

5.3. RQ3: Polysyllabic shortening in the foot nucleus of trisyllables as compared to disyllables (STUDY 2)

The influence of the number of syllables in the foot (2 vs. 3 syllables) on the duration and length of the foot nucleus elements is explored in structures with the four shortest foot nucleus types. Duration of V_1, C_2 and V_2 is assessed as a function of an interaction of the foot nucleus structure and a number of syllables in the foot: [V_1 / C_2 / V_2] duration ~ Foot nucleus structure * Number of syllables in the foot + (1 | speaker) + (1 | wordform), data = 812 tokens.[7]

This interaction is insignificant for V_1 ($\chi^2(3)$ = 1.08, n.s.), significant for C_2 ($\chi^2(3)$ = 14.28**, p = 0.003), and highly significant for V_2 ($\chi^2(3)$ = 26.76***). In other words, in the shortest disyllables and trisyllables, V_1 duration does not depend on the overall length or structure of the foot, while the duration of C_2 and V_2 does. Modelled means and relevant post-hoc pairwise comparisons between them in C_2 and V_2 are given in Appendix 5. Below we highlight the most important findings

[6] In this pair, the difference between estimated least squares means is considerably larger than between observed raw means between the VCV and VC·V: nuclei (12 ms in Fig. 3) due to the presence of V_2 phonological length in the model. The difference still remains significant if "V_2 etymological length" is excluded from the model both as random and a fixed effect (13 ms**, p = 0.006), or if a model estimates V_2 duration as a function of the whole foot nucleus structure ("Foot nucleus structure"), where the predictor levels exactly correspond to the individual boxes in Appendix 4a (14 ms*, p = 0.024*): V_2 duration ~ Foot nucleus structure + (1 + V_2 length | speaker) + (1 | wordform) + (1 | V_2 quality) + (1 | C_2 quality). Kuznetsova and Markus (2022) showed that the degree of a difference between V_2 in VCV vs. VC·V: varies across speakers, but the speaker-specific random slope either for "Foot nucleus structure" or for the ternary interaction "V_1 length * Presence of R * C_2 length could not be included in any of these overall models on V_2 due to the model convergence failure.

[7] The formula for V_1 additionally included the random intercepts for C_1 presence / absence and for C_1 quality, which also both resulted significant.

pertinent to our hypothesis **H3** about polysyllabic shortening in the trisyllabic foot as compared to the disyllabic one. First, we address the sub-hypotheses 1 and 2 about the differences between disyllables and trisyllables in the structure of the V_2 length contrast (Section 5.3.1) and in the strength of the enhancement of the ternary quantity contrast with the durations of V_1 and V_2 (Section 5.3.2). Then we briefly look at the situation in the ternary quantity contrast itself (Section 5.3.3).

5.3.1. Sub-hypothesis 1 of H3: Difference in V_2 duration and quantity between disyllables and trisyllables

Fig. 4 shows the raw (4a) and modelled means (4b) of V_2 in the four shortest disyllables and trisyllables. Our general hypothesis **H3** about shorter durations in trisyllables than in disyllables is supported. However, we also see that the degree of V_2 shortening in trisyllables depends on the foot nucleus structure.

The least difference between disyllables and trisyllables is observed for the "half-long" V_2 in the VCV nucleus of the light foot (see **H2A**). The modelled distance between V_2 durations in the disyllabic vs. trisyllabic light foot (*tap**a*** – *lak**a**ta*; 16 ms*, p = 0.013, cf. the first black and grey points from the left in 4b and Appendix 5) is much smaller than for V_2 durations in the three studied types of heavy disyllabic vs. trisyllabic foot.

In the heavy foot, original long V_2 in trisyllables with the VC:V: nucleus (the second box from the right in 4a) is now short as part of the general long vowel shortening process discussed in Section 5.2. In our model, it differs from the original short V_2 in the VC:V nucleus of trisyllables only by 13.5 ms (**kattīma* > *katt**i**ma* – *katt**i**la*; n.s.; cf. the first two grey points from the right in 4b), which is far below the audibility threshold.

As predicted by **sub-hypothesis 1** of H3, disyllables differ from trisyllables in that they do not exhibit phonological shortening of long V_2. In particular, the studied disyllables with the VC:V: nucleus retain a phonologically long vowel (the third box/black point from the left in 4a-b). The modelled difference between long and short V_2 in the VC:V: – VC:V pair of disyllables (*tapp**ā*** – *natt**a***; the first two black points from the right in 4b) is 40 ms***. As a result of the phonological shortening of long V_2 in trisyllables, the modelled difference in the V_2 duration of the VC:V: nucleus in trisyllables vs. disyllables (**kattīma* > *katt**i**ma* – *tapp**ā***) is very large, -64 ms***. The shift of the originally long V_2 in the VC:V: nucleus of a trisyllable into the class of short vowels vs. the lack of this shift in disyllables is marked as "V:" vs. "V" under the boxes in 4a.

Phonetic polysyllabic V_2 shortening in the heavy foot is signalled by a difference of -32 ms*** in the V_2 duration between the disyllabic and trisyllabic feet with the VC·V: nucleus (*mať**ť**āla* – *taǒ**p**ā*) and of -38 ms*** for the VC:V nucleus (*katt**i**la* – *natt**a***). These differences are in between the minimal one found in the light foot and the maximal one found in the case of phonological shortening of long V_2 in a trisyllable.

5.3.2. Sub-hypothesis 2 of H3: Inverse relation of V_2 duration to C_2 ternary length

Here we investigate the **sub-hypothesis 2** of **H3**: the expected difference between disyllables and trisyllables regarding the enhancement of the ternary length contrast in C_2 by the inversely related duration of V_2.

The most comparable triplet of disyllables (*tap**a*** — *tap̆p**ā*** — *tapp**ā**;* cf. the first three black points from the left in 4b and detailed data in Appendix 5) manifests only a slight compensatory trend in V_2, as also predicted by Markus (2011). In VCV, the modelled mean duration of V_2 is 144 ms, in VC·V: it is 139 ms, and in VC:V: it is 131 ms. The difference in modelled V_2 duration is weakly significant between the first and third members of this triplet (p = 0.032*), but insignificant between the first and second members and the second and third members.

In the corresponding triplet of trisyllables (*lak**a**ta* – *maťť**ā**la* – **kattīma* > *katt**i**ma*), the modelled means of V_2 duration are 128, 107, and 66 ms, respectively. Importantly, all three mutual differences are highly significant (cf. the first three grey points from the left in 4b and Appendix 5). This finding shows that the shortest disyllables and trisyllables differ in the degree of development of the inverse relation between V_2 duration and C_2 length (see discussion in Section 6.2).

Modelled V_1 means in this triplet barely differ, both between disyllables and trisyllables, and across the foot nucleus types. Pairwise comparisons between the modelled least squares means across all eight structures show insignificant differences in a range of 0–1.5 ms in disyllables and of 1.3–7.5 ms in trisyllables. Modelled V_1 durations in trisyllables are slightly shorter than in disyllables (by 0.5–6.5 ms), which is generally in line with hypothesis H3, but no clear compensatory trend can be observed.

5.3.3. Hypothesis H3 vs. anticompensatory lengthening in C_2

Fig. 5 illustrates raw and modelled C_2 durations in the four disyllables and four trisyllables of STUDY 2. These data confirm a robust ternary quantity contrast in the C_2 of both disyllables and trisyllables (cf. modelled 86 vs. 205 vs. 271 ms*** and 89 vs. 187 vs. 237 ms*** in the best comparable nuclei VCV — VC·V: — VC:V: in Appendix 5). Smaller ratios in trisyllables (1 : 2.1 : 2.7) than in disyllables (1 : 2.4 : 3.2) are primarily due to the shorter durations of both types of geminates in trisyllables.

Hypothesis **H3** predicts shorter durations of segments in trisyllables than in disyllables, which we indeed observe in Fig. 5 as a general trend. However, H3 does not explain why significantly shorter durations are observed in the C_2 of trisyllables only before phonologically long or shortened long vowels (-18 ms*, p = 0.043, in VC·V:, *ma**ťť**āla* – *ta**p̆p**ā*; -34 ms*** in VC:V:, **kattīma* > *ka**tt**ima* – *ta**pp**ā*). Before phonologically short vowels, neither a singleton nor a long geminate shows any significant differences between disyllables and trisyllables (*la**k**ata* – *tapa* and *ka**tt**ila* – *na**tt**a*; the first and the last pair of black and grey points in 5b). In contrast to C_2, the durations of V_2 in trisyllables are always significantly shorter than those in respective disyllables (see Section 5.3.1).

There is, therefore, a discrepancy between coherently longer V2 but incoherently longer C_2 in disyllables than in trisyllables. We interpret this discrepancy through anticompensatory lengthening of C_2 which is present only before phonologically long vowels (see further in Section 6.4). The most prominent anticompensatory effect in Finnic varieties is phonetic lengthening of both short and long consonants before long vowels, and secondary gemination is a phonologised outcome of such lengthening (Kuznetsova, accepted). The pro-

(a) Raw mean durations of V_2

(b) Modelled influence of the number of syllables in the foot and the foot nucleus structure on V_2 duration

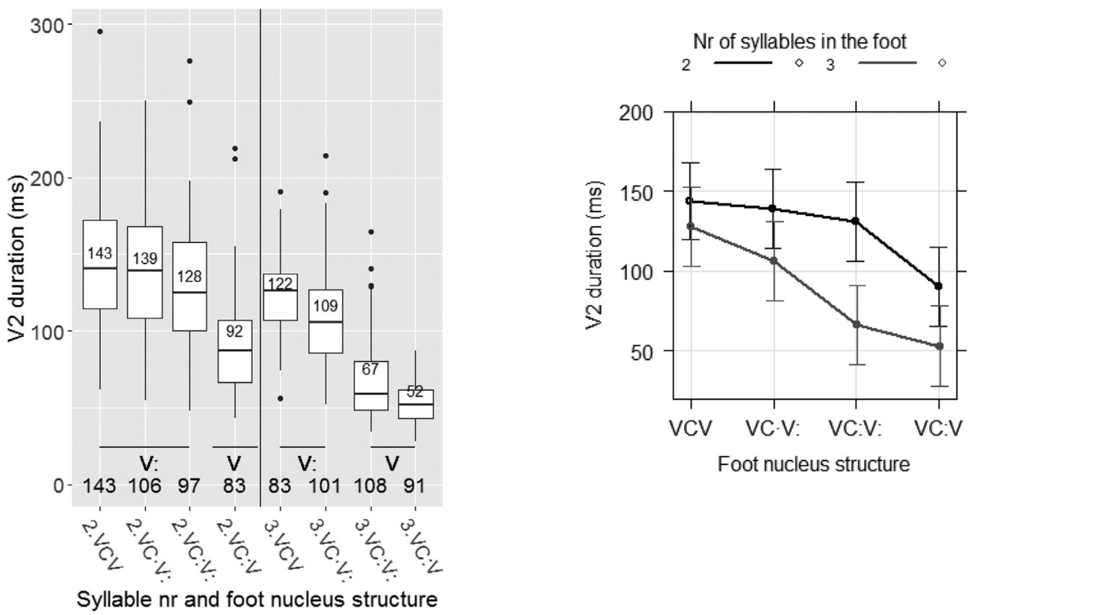


Fig. 4. Duration (in ms) of V_2 in disyllables vs. trisyllables with the four shortest foot nuclei. Boxes are coded by the foot nucleus types reported in Table 2.

(a) Raw mean durations of C_2

(b) Modelled influence of the number of syllables in the foot and the foot nucleus structure on C_2 duration

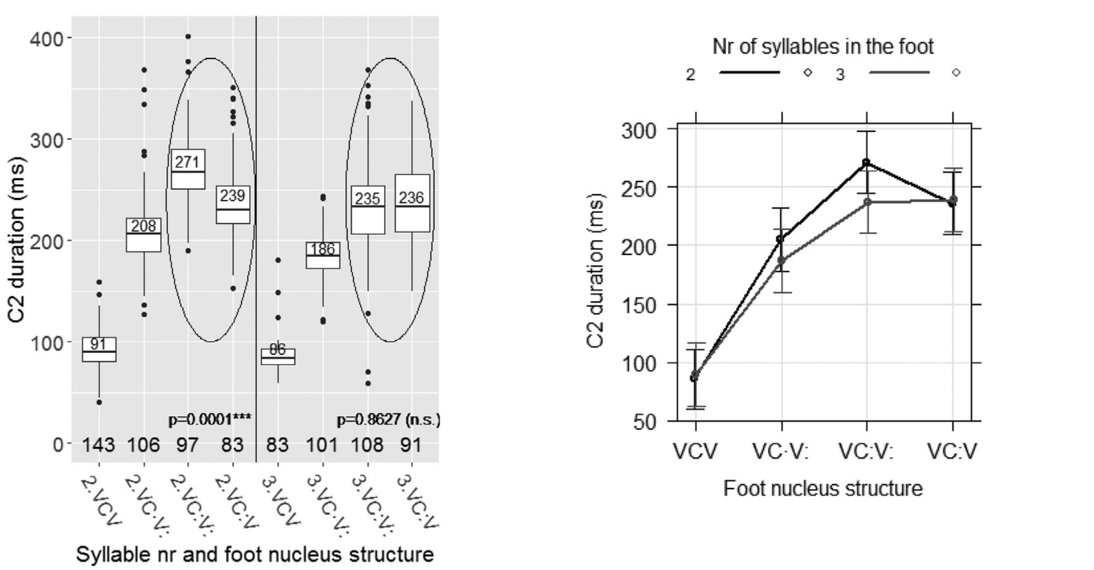


Fig. 5. Duration of C_2 in disyllables vs. trisyllables with the four shortest foot nuclei (cf. legend to Fig. 4).

cess can also act in the opposite direction: if a long vowel shortens, phonetic lengthening of the preceding consonant can disappear (Palander, 1987, p. 230).

In our foot nuclei VC·V: and VC:V:, we can see a clear direct correlation between the real synchronic duration of originally long V_2 and the degree of anticompensatory lengthening in preceding short and long geminates (cf. also Appendix 5). In disyllables vs. trisyllables with the VC·V: nucleus (*mat̆tāla* – *tap̆pā*), the difference between the two modelled long vowels in V_2 is 32 ms*** and between the two preceding consonants in C_2 is 18 ms* (p = 0.043). In feet with the VC:V: nucleus (**kattīma* > *kattima* – *tappā*), the difference between disyllables vs. trisyllables is much larger for both V_2 and C_2. For originally long V_2, it is 64 ms*** (and reflects the difference between truly long and shortened long vowels, cf. Section 5.3.1), and for C_2, it is 34 ms***.

The C_2 long geminates which precede shortened long vowels in the *VC:V: nucleus of trisyllables do not undergo anticompensatory lengthening anymore. Their modelled duration is equal to that of the C_2 before originally short vowels in the VC:V nucleus of trisyllables (**kattīma* > *kattima* – *kattila*; cf. the first two structures from the right outlined by an ellipsis in 5a). Disyllables, however, retain phonologically long vowels after the primary geminates in the VC:V: nucleus: the modelled difference between long and short V_2 in the VC:V: — VC:V pair (*tappā* – *natta*) is 40 ms*** (see Section 5.3.1). Consequently, there is a very significant durational difference between preceding consonants (modelled 35 ms***, p = 0.0001; cf. also the 3rd and the 4th structure from the left outlined by an ellipsis in 5a).

In sum, long and short geminates undergo phonetic anticompensatory lengthening before synchronically phonologically long vowels in our data: the longer the vowels are, the longer the preceding consonants are. For phonologically short vowels, including the "half-long" V_2 in the light foot, such a direct correlation is not observed: while these V_2 are longer in disyllables than in trisyllables, the preceding C_2 consonants are not.

5.4. RQ4: Polysegmental shortening in the elements preceding V_2 in trisyllables (V_1, C_2, R in STUDY 1)

Hypotheses regarding the structure-related phonetic shortening of foot nucleus segments preceding V_2 predict that long phonemes will manifest significant compensatory effects (**H4A**), while short phonemes will show weak or no effects (**H4B**). Expected effects are linked to compensatory shortening under the influence of greater quantity of neighbouring V_1 and C_2 and the presence of a sonorant R. Secondary geminates are expected to behave as long phonemes, i.e., like primary (long) geminates rather than like singletons.

These expectations are generally confirmed. In many cases, short members of the length categories of both V_1 and C_2 show no statistical differences, while longer members do. However, sometimes short phonemes (short V_1, singleton C_2, and especially sonorant R) manifest significant compensatory effects in trisyllables, although at most weak effect was expected based on the earlier results on disyllables.

The following formulas are used to model the duration of V_1, C_2, and R:

V_1 duration ~ V_1 length * presence of R + V_1 length * C_2 length + V_1 length * presence of C_1[8] + (1 | speaker) + (1 | wordform) + (1 | C_1 quality) + (1 | V_1 quality), data = 3812 tokens;

C_2 duration ~ C_2 length * presence of R + C_2 length * V_1 length + C_2 length * diphthong[9] + (1 | speaker) + (1 | wordform) + (1 | V_2 quality), data = 3812 tokens;

R duration ~ V_1 length + C_2 length + (1 | speaker) + (1 | wordform) + (1 | R quality), data = 1399 tokens.

5.4.1. Mutual influence of V_1 and C_2 length on each other's durations

A robust finding in earlier studies of Finnic languages, including those on Soikkola Ingrian, was that increased quantity in V_1 and C_2 shortens the durations of long C_2 and V_1 respectively (**sub-hypothesis 1** of **H4A**). Short V_1 and singleton C_2 either did not change their duration significantly in the same conditions or showed a weak compensatory trend (part of **H4B**). Our results presented in Fig. 6 confirm these expectations, with one correction concerning short vowels before singleton consonants.

First, the longer the following consonant, the shorter the V_1 duration (6a). Modelled long vowels before secondary (short) geminates are 18 ms*** shorter than before singletons, and 11 ms** (p = 0.002) longer than before primary (long) geminates. The inverse relation of long V_1 duration to the ternary length of the following C_2 is significant in all cases, as expected. This relation is an additional cue reinforcing ternary length in structures with a long vowel or a diphthong as V_1.

For short V_1, weak or no compensatory effects are expected. In our data, modelled short vowels before secondary geminates are 16 ms*** shorter than before singletons and just 4 ms (n.s.) longer than before primary geminates[10]. In short vowels before singleton consonants, therefore, the discussed trend for the inverse durational relation is highly significant instead of weak or absent, i.e., stronger than expected.

Second, the longer the preceding vowel, the shorter the duration of any geminate C_2 (6b), while singletons are not significantly affected in a compensatory way (3 ms; n.s.). Modelled secondary geminates are 25 ms*** longer, and primary geminates are 21 ms*** longer before short vowels than before long vowels. This result entirely corresponds to our predictions.

5.4.2. Influence of the presence of a sonorant R on V_1 and C_2 durations

Sub-hypothesis 2 of **H4A** predicts that a presence of an additional segment, in our case the sonorant R, in the first syllable will shorten the durations of both long V_1 and geminate C_2. Hypothesis **H4B** says that short V_1 and C_2 will not be significantly affected in such a case.

These expectations are confirmed by our data for C_2, with a correction (Fig. 7). Instead of only long phonemes, all the three length classes of C_2 are shortened if preceded by a sonorant: modelled singletons by 19 ms** (p = 0.002), secondary geminates by 29 ms***, and primary geminates by 34 ms***. Singletons, therefore, are also significantly shortened in the presence of R, rather than manifesting weak or no effects, i.e., the effect is stronger than expected. Interaction between the length class of C_2 and the presence of a sonorant is insignificant. This shows that there are no big differences in the degree of durational shortening of

[8] A trivial shortening effect of the presence of a word-initial consonant C_1 on V_1 duration, significant but not discussed here.

[9] The factor of whether a long V_1 was a diphthong or a monophthong, which showed a very weak anticompensatory effect (significant only for modelled long geminates, which differed by 19 ms***), not discussed here.

[10] Slightly shorter V_1 duration before short geminates than before long geminates is likely because the former group includes structures with the longest type of the first syllable, V:RC·, while the latter does not, due to their absence in Ingrian; see Table 2 and Appendix 1.

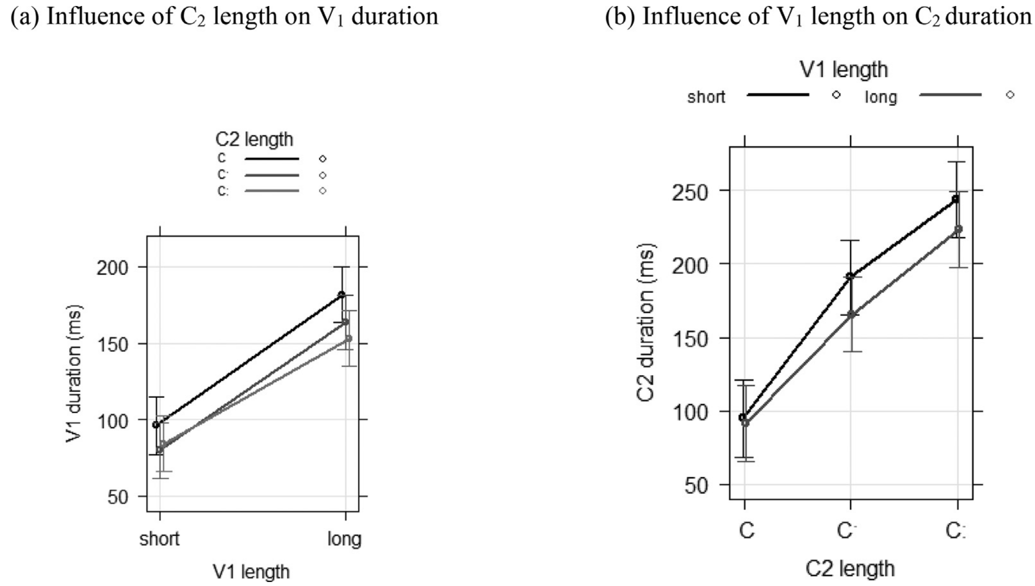


Fig. 6. Modelled mutual influence of V_1 and C_2 length on each other's durations.

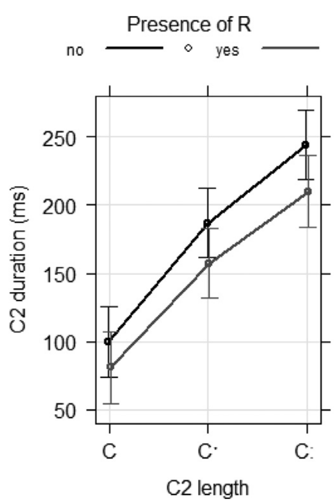


Fig. 7. Modelled influence of the presence of R on C_2 duration (interaction is n.s.).

the three length classes in the presence of R. Still, both types of geminates are shortened a bit stronger than singletons[11].

5.4.3. Influence of V_1 and C_2 length on the duration of a sonorant R

The properties of a sonorant R itself have not been explored well. Scarce Estonian data indicated a trend for a shorter duration of R in the proximity of longer length classes of V_1 and C_2. As the sonorant R is a short phoneme, we suggest a weak or no effect for it in our hypothesis **H4B**.

Our results (Fig. 8), however, show a significant inverse relation of the sonorant duration to the length of both V_1 and C_2 (raw means of R in respective structures are also given in Appendix 4d). Long V_1 shorten modelled R by 11 ms*** (8a). Before a singleton C_2, R is 19 ms*** longer, and before a long geminate, R is 11 ms*** shorter than before a short geminate (8b).

It can be said that the effect of an inverse relation between the duration of R and the length of C_2 also reinforces the ternary quantity contrast of consonants in respective Soikkola Ingrian trisyllables, much like the inverse relations between the C_2 length and the durations of V_2 and long V_1. An inverse relation of the duration of R to V_1 length enhances also the binary length contrast of V_1 in respective structures.

6. General discussion

6.1. Synthesis of main findings

This section provides an overview of our main findings compared with our original expectations. We further elaborate on the theoretically relevant issues in Section 6.2 (ternary quantity contrast), Section 6.3 ("half-long" vowel), and Section 6.4 (relation of compensatory effects to foot isochrony and stress).

Aggregated raw mean durations of all foot segments (including the third syllable, but excluding C_1) in both studies are given in Appendix 3. Our results on the compensatory effects of foot structure in the durations of foot nucleus elements in trisyllables are generally in line with previous findings. They also clarify certain aspects of our hypotheses and highlight some expected and newly discovered differences between disyllables and trisyllables regarding the scope and strength of compensatory effects.

The duration of the foot nucleus elements varies greatly as a function of foot structure. As in earlier work, the scope of variability progressively increases in greater length classes of segments. Phonological length contrasts are nevertheless always robust in V_1 and C_2 in structurally comparable contexts. Importantly, our study shows that the ternary length distinction is well-maintained in C_2 not only in the shortest trisyllables, as suggested by earlier observations, but also in longer trisyllabic feet.

[11] As for V_1, it shows a weak anticompensatory effect, being slightly longer before a sonorant R than directly before a C_2 stop (interaction n.s.). The effect is significant only for modelled short V_1 (10 ms**, p = 0.003) and is not addressed in this paper.

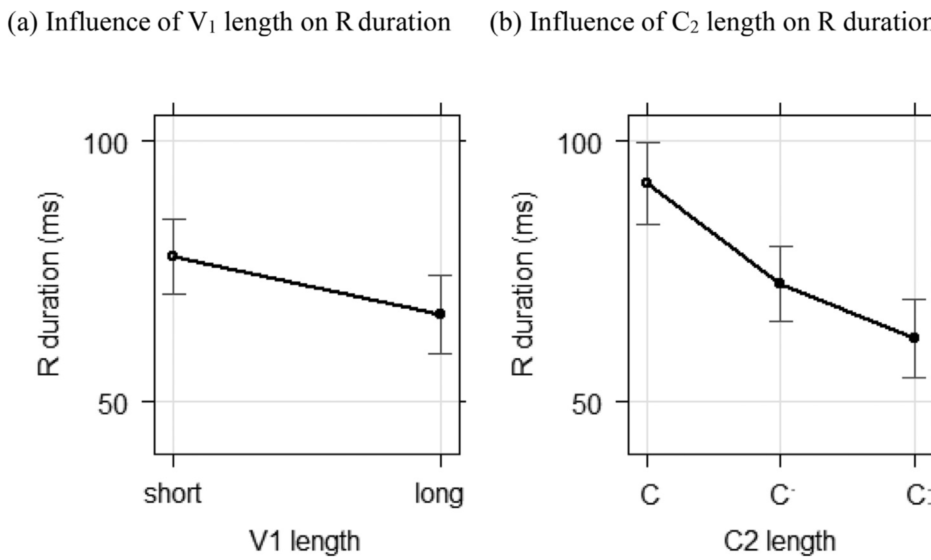


Fig. 8. Modelled influence of V_1 and C_2 length on the duration of a sonorant R.

In trisyllables, the overall ratio between the short and the long vowels is 1 : 2, and between the three consonantal length types is 1 : 2 : 2.5 (see also Table 5). Short vowels are similar in mean duration to singleton consonants and long vowels are similar in mean duration to short geminates, while long geminates can be characterised as phonetically overlong.

We observe many kinds of compensatory structural influences on segmental durations. Crucially, in trisyllables, these influences are more numerous and stronger than originally expected, especially for short phonemes. Compensatory shortening in trisyllables is most evident in long V_1, both geminate types of C_2, sonorant R, and long V_2 after the heavy syllable.

In the shortest foot nucleus types, the ternary quantity contrast is enhanced only by the inverse relation between V_2 duration and C_2 length, insignificant in disyllables and significant in trisyllables. In the shortest disyllables, the ternary quantity contrast is present in its most independent, "pure" form, which confirms earlier studies.

However, in more complex structures, important compensatory effects have been observed in earlier studies also in the first syllable. The only such effect explicitly discussed for the Soikkola Ingrian disyllables, as well as for Estonian and Finnish, has been a strong mutual compensatory shortening of long V_1 and the following short or long geminate C_2. This effect is entirely confirmed by our data.

On the other hand, at most weak compensatory effects have been generally observed in earlier studies for short phonemes within the foot nucleus: short V_1, singleton C_2, sonorant R (although the properties of the sonorant R have been little studied). In our study on trisyllables, short phonemes show significant compensatory shortening in several cases. We observe significantly shorter duration of short V_1 before geminates than before a singleton C_2 and a significant inverse relation of the duration of a sonorant R to an increase in length of both V_1 and C_2. The presence of a sonorant also significantly shortens the duration of the following C_2 of all quantity classes, including singletons. Soikkola Ingrian trisyllables, therefore, manifest stronger compensatory effects than those reported for disyllables not only in the duration and length of V_2, but also in the durations of preceding elements. The general trend is that the more complex the foot structure, the more numerous and stronger compensatory effects are observed in it. Consequently, in progressively more complex foot structures, the ternary quantity contrast of consonants receives progressively more reinforcement from the inverse relation between C_2 length and the durations of neighboring segments.

Most importantly, there is a significant difference between disyllables and trisyllables regarding the phonological contrast in V_2, which also correlates to the degree of reinforcement of C_2 ternary length by inversely related V_2 duration. In

Table 5
Durational ratios between length classes in C_2 in selected disyllabic and trisyllabic feet of Soikkola Ingrian, Estonian, and Inari Saami. For C_2, only the position after a short V_1 before a short or long V_2 is represented (i.e. the V_V(:) context). Soikkola Ingrian ratios showing the biggest differences from those in the listed cognate languages are in bold.

Foot type:		Disyllables			Trisyllables			
Type of contrast:		C: - C	C· - C	C: - C·	C: - C	C· - C	C: - C·	Source
Soikkola Ingrian	VC:V	2.7		**1.2**	2.7		**1.3**	Study 2
	VC:V:	3.2	**2.4**	**1.3**	2.7	**2.1**	**1.3**	
Estonian		3.5	2.1	1.7	3.2	1.9	1.6	(Lehiste, 1997, pp. 151, 158)
Inari Saami		2.8	1.5	1.8	2.3	1.5	1.6	(Türk et al., 2019, 2020)

disyllables, there is a slight isochronic trend towards the inverse relation of V_2 duration to C_2 length. Trisyllables represent a more innovative prosodic stage where two important changes have happened. First, the inverse relation of V_2 duration to C_2 length has become highly significant. Second, etymologically long V_2 in the heavy foot have shortened in all the trisyllabic structures but one. As an outcome, the quantity contrast of V_2 in trisyllables can be considered entirely lost, for two main reasons.

First, durationally long types of V_2 are attested only in two types of foot nuclei, VCV and VC·Vː. In these nuclei, there has been no V_2 length contrast historically (see Tables 2-3 and Appendix 1). Moreover, V_2 in these two types of nuclei cannot be combined into a single type of synchronic phonologically long V_2, because "half-long" V_2 in VCV is a phonetically lengthened short vowel, while V_2 in VC·Vː can indeed be considered phonologically long in synchrony (see Section 6.2).

Second, the modelled difference of 8.5 ms between etymologically short and long vowels is not significant in our model and is also far below the audibility threshold (expected to be at least 30–40 ms for length contrasts). The effect, therefore, cannot be classified even as a just-noticeable difference at the group level, although a detailed analysis of between-speaker variation (Kuznetsova & Markus, 2022) showed a complete merger in three out of our five speakers and an incomplete merger in two speakers (AG and EI), regulated in a compensatory manner by the complexity of the first syllable. Additionally, the modelled distinction between the original phonologically short and long vowels is in itself relatively artificial because it groups together Type 1 (phonetically short) & Type 4 ("half-long", i.e., phonetically lengthened) vowels, as opposed to Type 2 & Type 3 vowels from Fig. 3. The overall result, therefore, qualifies as a loss of the quantity contrast from the V_2 position of trisyllables.

In disyllables, long and short V_2 are still contrasted after long (primary) geminates and singletons. After short (secondary) geminates, the V_2 quantity contrast has never existed in disyllables. With its loss from trisyllables, the contrast in V_2 is now gone from Soikkola Ingrian altogether. A synchronic outcome of the structural phonological change in V_2 quantity, which can be compared to the original situation reflected in Tables 2-3, is presented in Table 4.

Although anticompensatory effects are not generally addressed in this paper, we have discussed the presence of phonetic anticompensatory lengthening of a long geminate C_2 before a long vowel in the VCːVː nucleus of disyllables, where the length contrast of V_2 is still preserved. In turn, the lack of such lengthening in a long geminate C_2 before a shortened long V_2 in a trisyllable with the *VCːVː nucleus further supports the phonological nature of long V_2 shortening in trisyllables.

Secondary (short) geminates, which originally emerged out of anticompensatory lengthening in singleton,s are now completely phonologised as long phonemes. With regard to compensatory effects, they phonetically behave like primary (long) geminates, i.e., as long phonemes rather than as singletons.

6.2. Soikkola Ingrian V_2 and C_2 quantity contrasts in the cross-linguistic context

Our main phonological finding, the difference in the structure of the V_2 length contrast between the Soikkola Ingrian disyllabic and trisyllabic feet, is typological in nature. As found in earlier research and confirmed by our study, in a triplet of the shortest disyllables (e.g. *kana* [ˈkanaː] 'hen' — *kaňnā* [ˈkan·aː] 'hen:PRT' — *linnā* [ˈlinːaː] 'city:PRT'), an inverse relation of V_2 duration to C_2 length is very weak, while short V_1 do not differ in their duration (in the case of stops and /s/, voicing of singletons is an additional phonetic cue, though; see Section 2). Such "pure" ternary quantity is cross-linguistically very rare, as ternary quantity contrasts of vowels and consonants are usually significantly enhanced by compensatory effects in the durations of adjacent segments (Remijsen & Gilley, 2008; McRobbie-Utasi, 2007; Kuznetsova, 2015). In particular, in other languages with consonantal ternary quantity, Estonian, Livonian, and Saami, this contrast in similar triplets (e.g. in Estonian *lina* [ˈlin**a**ː] 'linen' — *linna* [ˈlin·**a(·)**] 'city:GEN' — *linna* [ˈlinː**ă**] 'city:PRT'), is enhanced by statistically significant compensatory modifications in the durations of the second and (in Saami) of the first syllable vowels (cf. comparative overviews in Markus et al., 2013; Kuznetsova, 2015).

Out of the languages with ternary quantity for which relevant phonetic studies exist, the situation in Soikkola Ingrian disyllables is most similar to that of Dinka and Shilluk. In these West Nilotic languages with three distinctive vowel quantities, an inverse relation of the duration of a coda consonant also only slightly enhances the duration of a preceding vowel (Remijsen & Gilley, 2008; Remijsen et al., 2019). In these languages, ternary quantity is observed in the vowels of monosyllabic structures, i.e., (C)VC — (C)V·C — (C)VːC, while in Soikkola Ingrian ternary quantity is attested in consonants (a rarer case) and only in polysyllabic words. In such cases as Dinka, Shilluk, and Soikkola Ingrian, a phonological interpretation of the ternary contrast at the segmental level might be plausible, even if such an account tends to be avoided in phonological theory (cf. overviews in Prehn, 2012; Kuznetsova, 2018b).

Our study shows that Soikkola Ingrian trisyllables are more prosodically innovative than disyllables. Soikkola Ingrian trisyllables show prosodic properties typologically similar to those observed in other languages with ternary consonantal quantity. Instead of a former phonological length contrast in V_2, there is now a statistically significant inverse relation of V_2 duration to C_2 length. We can say that the factor of foot structure has completely taken over the factor of etymological length in regulating the duration of V_2 in Soikkola Ingrian trisyllables. Trisyllabic vowel shortening in Soikkola Ingrian can also be compared to trisyllabic vowel shortening in Old and Modern English (where it concerns stressed syllables, viz. Lahiri & Fikkert, 1999). The processes are quite different, but the very factor of the trisyllabic foot as a condition for vowel shortening is similar.

Coming back to the cognate languages, it is also noteworthy that in Soikkola Ingrian, Estonian, and Livonian, all synchronic durationally long V_2 classes originate from Proto-Finnic short vowels, while original phonologically long non-initial vowels have shortened or even reduced (see Section 6.3 on Soikkola Ingrian). This fact apparently reflects the opposite vectors of prosodic development in the light vs. the heavy foot: second syllable vowel lengthening vs. shortening. The nature of this relation of V_2 duration to the weight of the first syllable is compensatory, and this issue is raised again in Section 6.4.

It is also remarkable that in synchrony all the three Finnic languages manifest a contrast of short and long geminates before phonologically short vowels in trisyllables, but in Soikkola Ingrian it has entirely different origins than in Livonian and Estonian. As mentioned in the introduction, the middle length class has emerged through the lengthening of singletons in Soikkola Ingrian, rather than through the shortening of long geminates, as in Estonian and Livonian. In the slightly more distant Saami languages, both processes are attested (Sammallahti, 1998).

Moreover, Soikkola, Hevaha, and Oredež Ingrian are the only Finnic varieties which have developed secondary geminates before phonologically short vowels, and this has happened only in trisyllables. Outside the trisyllabic foot, secondary gemination exists only before long vowels (as in **kanā* > *ka**ňn**ā* ‘hen:PRT’) and certain diphthongs. In the Soikkola Ingrian trisyllabic foot, secondary geminates have also developed before a sequence of two light syllables *-CVCV(C): **murkina* > *mur**ǩk**ina* [ˈmurkˑina] ‘breakfast’, i.e. before a short vowel (see Kuznetsova, accepted).

As noted in Section 1, durational ratios in Finnic and Saami languages have been found to be more stable than absolute segmental durations. Table 5 reports a comparison of Soikkola Ingrian ratios between the modelled means of the length classes in the C_2 consonants of disyllables and trisyllables with the respective ratios in similar structures of Estonian and Inari Saami. It shows that Soikkola Ingrian short geminates are further from singletons and closer to long geminates than in the other two languages. In those Ingrian varieties where the ternary contrast has not been preserved, short geminates, indeed, have generally merged with long geminates (Laanest, 1987; Kuznetsova, 2015).

However, in Inari Saami, cited here due to the availability of data on trisyllables, short geminates stay the closest to singletons out of all acoustically studied Saami languages. The ratios in North and Lule Saami disyllables are actually more similar to those observed in Soikkola Ingrian (cf. Türk et al., 2019, p. 37). As for Soikkola Ingrian vs. Estonian, the differences between the ratios reported in Table 5 can indeed be hypothetically linked to the different origins of the middle length class in the two languages, in the sense that they have developed through a vector of lengthening in Soikkola Ingrian and through a vector of shortening in Estonian.

6.3. The nature of the “half-long” vowel

An important phonological difference between Soikkola Ingrian and the other two Finnic languages with a ternary quantity contrast in consonants, Estonian and Livonian, is that Soikkola Ingrian still maintains the contrast between long and short V_2 in its phonological system. In an Estonian disyllabic triplet with three degrees of quantity like *lina* [ˈlinaː] ‘linen’ — *linna* [ˈlinˑa(ˑ)] ‘city:GEN’— *linna* [ˈlinːă] ‘city:PART’, the duration of V_2 varies a lot but the vowel is always phonologically short, as there is no length contrast in unstressed vowels.

A comparable Soikkola Ingrian triplet *kana* [ˈkanaː] ‘hen’ — *kaňnā* [ˈkanˑaː] ‘hen:PRT’ — *linnā* [ˈlinːaː] ‘city:PRT’, on the contrary, exemplifies exceptional issues in mapping between the phonetic duration and the phonological quantity of unstressed vowels. Most notably, phonologically short V_2 in the VCV nucleus (the “half-long” vowel) in Soikkola Ingrian tends to have longer duration than any kind of phonologically long V_2. This situation is typical also for adjacent Finnic varieties (Votic, Ingrian Finnish, Karelian Finnish) and has created concerns in synchronic phonological descriptions of V_2 in these languages (see overview in Kuznetsova, 2016).

In Standard Finnish, the “half-long” vowel still contrasts with a true long vowel in the light foot: *tap**a*** ‘habit’ vs. *tap**aa*** ‘meet:3SG’. In the aforementioned innovative varieties, including Soikkola Ingrian, this contrast has been lost. However, the “half-long” vowel is still hardly ever considered as phonologically long there. Suomi and Ylitalo (2004) argued that the “half-long” vowel in Standard Finnish is part of stress manifestation, and we believe that this is also the case for Soikkola Ingrian and other Finnic varieties — more specifically, that the “half-long” vowel signals stress in the light foot. While a full-fledged discussion is out of place here, we will provide some arguments pertinent to the Soikkola Ingrian data.

First, this is a case of vowel lengthening in an unstressed syllable (a typologically rare feature). The foot-initial syllable is considered to be stressed not only in the heavy foot, but also in the light one, because this syllable serves as an anchor for high pitch. As noted in Section 2, high pitch might signal prominence at lexical or postlexical levels rather than at the foot level, but for the present discussion this is not important. Essential is the very fact that it is always the foot-initial syllable which gets high pitch. In our dataset of phrase-final tokens, the average pitch curve was also falling throughout the word from the first to the third vowel, with a short slightly rising plateau in the end of the final vowel (apparently signalling a phrasal boundary tone). This is in line with previous findings on stress-related pitch in Finnic languages.

Second, the positions of “half-long” vowels in multifoot words depend on the stress pattern used. As mentioned in Section 2, stress placement can be variable for rhythmic stresses, and in this case the positions of “half-long” vowels also vary, cf. Estonian *´kavalamalegi* [ˈkav**a**ˑˌłam**a**ˑˌleg̊**i**ˑ ∼ ˈkav**a**ˑła ˌmaleˑg̊i] ‘cunning:CMP:ALL:EMPH’, i.e. ‘even to a more cunning [e.g. person]’ (Eek & Meister, 2004; Lippus, Pajusalu, & Teras, 2006). In Soikkola Ingrian, where secondary gemination occurs also in the secondary stressed disyllabic and trisyllabic foot (Sovijärvi, 1944, pp. 12–15), variable stress placement can produce not only variable “half-long” vowels, but also variable secondary gemination: *´opastajoja* [ˈob̥aˑʃtăˌjojaˑ] ∼ *´opastaǰjōja* [ˈob̥aˑʃˌta**j**ˑoːjă] ‘teacher:PL:PRT’.

Third, the degree of V_2 lengthening in the light foot in Finnic languages is a function of the language variety and even the particular speaker, the degree of stress, as well as of the foot structure and the word structure (in longer structures, there is less lengthening) (Wiik & Lehiste, 1968; Lehtonen, 1970; Lehtonen & Leskinen, 1973; Leskinen, 1978; Pajusalu et al., 2005; Lippus et al., 2006). In Soikkola Ingrian, the degree of V_2 lengthening can also distinguish between primary and secondary stressed light feet (Kuznecova, 2009b, p. 32). Secondary stressed feet can also entirely lose stress in speech. Consider a Soikkola Ingrian example with two consecutive pronunciations of a bifoot word *´juma´likas* [ˈjum**a**ˑˌłig̊**a**ˑʃ] ‘saint’ in Fig. 9. In the first token under a phrase-final broad accent, V_2 is longer than V_1 in both feet. In the second token, pronounced in an unaccented phrase-initial position, foot stress (and so the

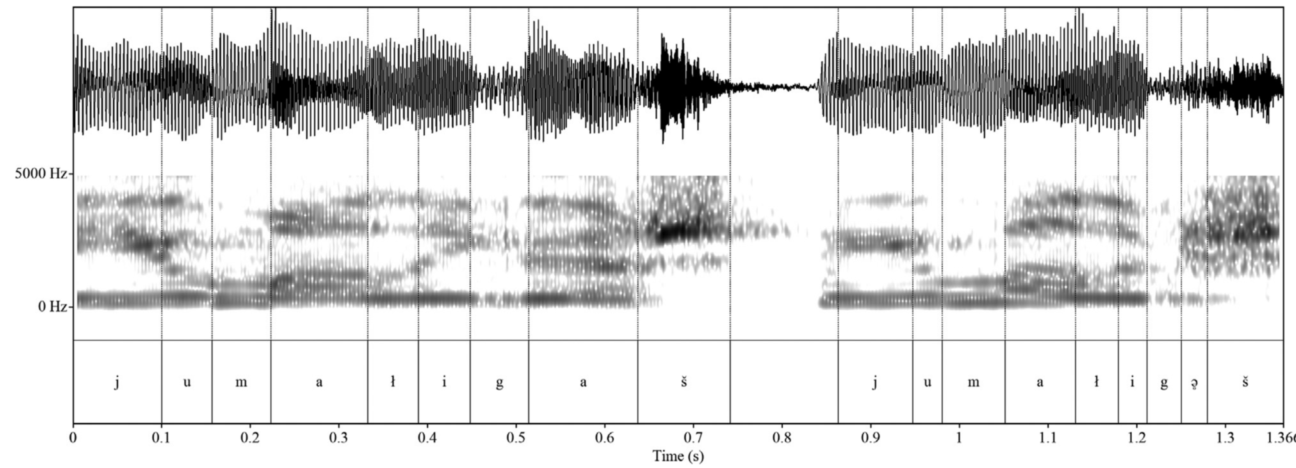


Fig. 9. Consecutive pronunciations of *jumalikas* [ˈjum**a**ˑˌɫig̊**a**ˑʃ] 'saint' under phrase-final stress and in an unaccented phrase-initial position (speaker VF, born 1935 in *Mäkkülä*, a southern Soikkola variety).

V_2 lengthening) is maintained in the primary stressed foot but not in the secondary stressed one. In the latter, V_2 is reduced to a voiceless schwa [ə̥] which is even slightly shorter than V_1, so the durational ratio between the two is entirely inverted.

The cases where the same vowel can be either "half-long" or reduced depending on footing or the degree of stress (like the final vowel in 'teacher:PL:PRT' and 'saint', respectively) and where the durational ratio between segments can be inverted are quite remarkable. They demonstrate quite clearly the non-segmental nature of V_2 lengthening in the light foot. As pointed out in the introduction, the durational ratios between segments tend to be well-preserved in Finnic languages in cases of true segmental length contrasts, because the ratios are the most important phonetic cues for these contrasts.

Note also that secondary gemination does not occur before the "half-long" vowel, unlike before true phonologically long vowels and diphthongs. In our study, singletons before the "half-long" vowel did not manifest any anticompensatory lengthening (a phonetic precursory for secondary gemination) either, unlike geminates before truly long vowels in disyllables (see Section 5.3.3).

An interesting exception in this respect is a synchronic VCˑV: nucleus in the Soikkola Ingrian trisyllabic foot, where the only durationally long variant of V_2, apart from the "half-long" vowel, is found in our study for trisyllables. As seen from Table 2 and Appendix 1, this nucleus does not contrast with the *VCˑV nucleus. There has been no contrast between long and short V_2 after a secondary geminate in this case, unlike other nuclei with secondary geminates in the trisyllabic foot. The VCˑV: nucleus in the Soikkola (and Hevaha and Oredež) Ingrian trisyllables is the only known Finnic case where secondary gemination is found before an original Proto-Finnic "half-long" vowel: **omena* [ˈom**e**:na] > *om̆mēna* 'apple'.

However, secondary gemination here is historically very late and explained by the general factor of the following sequence of two light syllables *–CVCV(C) in the trisyllabic foot (cf. Section 6.2 and Kuznetsova, accepted) rather than by a "half-long" vowel *per se*. It is still remarkable that the phonetic lengthening of originally short V_2 has been phonologised in this case. The phonologisation of secondary gemination in this nucleus has changed the structure of the first syllable from light to heavy. The vowel length contrast in the heavy foot synchronically exists in the language system, unlike in the light foot. Therefore, in Soikkola Ingrian, V_2 in a given position cannot be considered as phonologically short. In our data, no stress-dependent distortion of the V_1 and V_2 durational ratio has been attested for the VCˑV: nucleus, which also confirms the synchronic phonological status of the long V_2 here, unlike in the VCV nucleus.

6.4. A possible relation of compensatory effects to foot isochrony and foot stress in Soikkola Ingrian

Compensatory effects have often been linked to a general tendency toward isochrony (especially foot isochrony) for Finnic languages (see e.g. Krull, 1999; Lippus et al., 2013 on Estonian; Lehiste et al., 2008 on Livonian; Markus, 2010 on Soikkola Ingrian). In the "stress-timed" (i.e. isochronic) languages, stress is supposed to occur in equal intervals irrespective of the size of the stress groups (Pike, 1945; Abercrombie, 1967). Foot isochrony implies that the general duration of a metrically stressed foot tends to be preserved irrespective of its particular structure. Stronger compensatory effects discovered in our study both in the second and the first syllable of trisyllables, as compared to disyllables, might potentially indicate a tendency toward foot isochrony.

Isochrony, however, is a controversial concept (Turk & Shattuck-Hufnagel, 2013; Tilsen & Arvaniti, 2013), and its metrics return very different results (Loukina et al., 2011; Arvaniti, 2012; Maddieson, 2018; Romano, 2020). For Standard Finnish, isochrony has not been confirmed as a general tendency (Lehtonen, 1970, p. 143; Suomi & Ylitalo, 2004; Suomi, 2009, p. 415), in part due to the existence of anticompensatory effects, which add more duration in connection with more length and, therefore, contradict isochrony.

We could not directly test the foot isochrony hypothesis on our data. Such a test would imply a demonstration that syllables (the immediate prosodic sub-units within the foot), are shorter within longer feet. Where our tokens contain geminates at syllable boundaries, it is impossible to define the exact syllable boundaries. Still, we can compare an increase in overall foot duration and overall foot length.

For an illustration, an overall increase in length and raw duration across the four disyllables and four trisyllables from

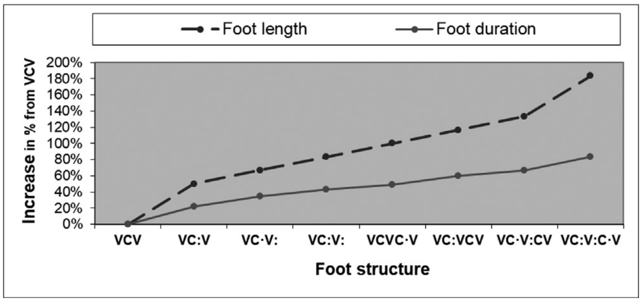


Fig. 10. Percentual increase in overall foot length and duration in the structures of STUDY 2 as compared to the VCV disyllable.

STUDY 2 is represented in Fig. 10 (C_1 was taken off all counts due to its absence in some words). The length increase is quantified in the following way. Approximate quantity index is ascribed to each segment in the eight structures from STUDY 2 based on the relative mean ratios of segmental durations that we have observed: 1 : 2 in vowels and 1 : 2 : 2.5 in consonants. Phonologically short vowels and consonants receive an index of 1, long vowels and secondary geminates, an index of 2, and overlong primary geminates, an index of 2.5. The overall length index is calculated for each of the eight structures as follows: VCV = 3, VC:V = 4.5, VC·V: = 5, VC:V: = 5.5, VCVC·V = 6[12], VC:VCV = 6.5, VC·V:CV = 7, VC:V:C·V = 8.5. Then we compare the percentual increase in this overall length across the eight structures with the percentual increase in their overall raw mean duration in relation to the shortest structure (the light disyllabic foot VCV).

Fig. 10 shows a non-linear relation between an increase in overall phonological length vs. an increase in overall duration across structures, which is expected of foot isochrony (duration increases much slower than length).

In our study, however, words were placed in phrase-final position under a broad phrasal accent, and the strength of a structural impact on duration also depends on prosodic context. Cross-linguistically, different prosodic positions at micro- (e.g. intrasyllabic) and macro- (e.g. phrasal or text) levels show stronger or weaker compression or lengthening effects (White, 2014; Rathcke & Smith, 2015). In particular, the phrase-final position is known to cause utterance-final lengthening (Lehiste, 1972; White, 2002). In our case, the second syllable of disyllables was word- and phrase-final, while the second syllable of trisyllables was not, because it was followed by the third syllable. Longer durations in the second syllable of disyllables could theoretically be explained by this fact.

Moreover, it has been shown that stressed vowels in words under phrase-final broad focus are especially susceptible not only to sentence-final lengthening (White, 2014), but also to polysyllabic shortening (White & Turk, 2010). If isochronic patterns are triggered by stress, they might be especially strong in our case, where foot stress is likely to be additionally strengthened by metrical prominence at higher prosodic levels.

In our study, all compensatory effects, including polysubconstituent shortening, were the strongest in the second syllable vowel. The "half-long" V_2 in the disyllabic and trisyllabic light foot vs. the phonologically shortened long V_2 in the heavy foot of trisyllables is probably the most prominent factor contributing to surface isochronic patterns in Soikkola Ingrian, as seen in Fig. 10 and Appendix 3, including the inverse relation of V_2 duration to the length of a preceding C_2 (cf. Section 5.3.2). As discussed in Section 6.3, the degree of phonetic V_2 lengthening in the light foot depends to a great extent on the degree of stress. Therefore, in a different phrasal position, the effect of foot isochrony might not be as strong as observed in this study because the "half-long" vowel might be much shorter (cf. the unaccented light foot in Fig. 9 discussed in Section 6.3).

Bearing these limiting factors in mind, let us summarise the interplay between phonological quantity and compensatory effects (which create an appearance of isochrony) observed in our study.

First, some phonetic studies suggest that languages with complex quantity strongly regulate speech tempo and phrasal stress effects in order to preserve their quantity systems from distortion, as e.g. in Finnish and Estonian (Lehtonen, 1970, p. 35; Krull, 1997; Eek & Meister, 2003; Nakai et al., 2009, 2012; Suomi et al., 2013; Plüschke, 2013). In particular, Suomi et al. (2003, 2013) observed little durational differences between unaccented and moderately accented words in Finnish and Estonian. Lengthening effects were significant only in strongly accented words, and our study lacked narrowly focused phrases with strong phrasal accent. In addition, as mentioned in the introduction, quantity languages keep essential durational ratios between segments relatively stable across various prosodic conditions. Therefore, no strong phrasal effects affecting segmental durations were necessarily expected for Soikkola Ingrian.

Second, let us comment in particular on possible phrase-final lengthening which can be present in the word-final second syllable of disyllables but not in the word-internal second syllable of trisyllables in STUDY 2. Earlier studies have shown greater magnitude of this effect in the rhymes of phrase-final syllables, as compared to onsets (Turk & Shattuck-Hufnagel, 2020, p. 33).

In STUDY 2, the duration of V_2 is in general a bit longer in disyllables than in trisyllables (Section 5.3.1), which can be at least partially attributed to possible phrase-final lengthening in disyllables. However, this fact can be likewise explained by general polysyllabic shortening in trisyllables as compared to disyllables, which, as noted, is especially typical for the phrase-final position. Most importantly, the phonological shortening of long V_2 in trisyllables, reported in earlier research and confirmed by our study, cannot be explained just by the lack of phrase-final lengthening in the second syllable of trisyllables, as compared to the second syllable of disyllables. On the contrary, the long V_2 shortening can be well accounted for by polysyllabic and polysegmental shortening in trisyllables.

Other foot nucleus elements in STUDY 2 do not show consistently longer durations in disyllables than in trisyllables. Phonologically short C_2 and V_1, as well as long geminates C_2 before the phonologically short vowels do not durationally differ between disyllables and trisyllables, while short and long geminates before originally long V_2 are longer in disyllables (cf. Section 5.3.2–5.3.3 and Appendix 5). Crucially, possible phrase-final lengthening in the second syllable of disyllables

[12] Here and in VC:V:C·V, C_3 is in all cases a long geminate shortened to a short geminate (see the full list of tokens in Supplementary material and Kuznetsova et al., 2023), so we gave it an index of 2.

vs. its lack in the second non-final syllable of trisyllables does not explain the discrepancy between the consistenly longer durations of V_2 but the inconsistently longer durations of C_2 in disyllables.

On the other hand, the observed difference in the duration of long geminates before originally long V_2 in disyllables vs. in trisyllables can be accounted for by anticompensatory lengthening before truly long vowels in disyllables vs. no lengthening before shortened long vowels in trisyllables (see Section 5.3.3). Importantly, in Suomi et al. (2013), anticompensatory lengthening was observed in Finnish under three different types of phrasal accentuation (unaccented, accented, contrastively accented), so this effect does not depend on phrasal accentuation. Note that the presence of living anticompensatory effects in Soikkola Ingrian indicates that "ideal" isochronic patterns are hardly found in this variety, as in Standard Finnish.

Third, let us consider an influence of compensatory effects on the system of phonological quantity in Soikkola Ingrian compared with earlier research. This variety has an extremely complex quantity system and, as noted above, is expected to regulate speech tempo and metrical prominence effects in order to keep the ratios related to quantity contrasts intact. Standard Finnish and Estonian, based on which such conclusion has been made, are not, however, varieties in the middle of sound change, contrary to Soikkola Ingrian. Estonian has apparently undergone a long V_2 shortening very similar to what is observed in Soikkola Ingrian. This ongoing process in Soikkola Ingrian can reveal us some of its internal mechanisms and dynamics, already obscure in Estonian. Long V_2 shortening in Soikkola Ingrian has a compensatory nature, because its degree is in an inverse relation both to the complexity of the first syllable and to the overall number of syllables in the foot (cf. also Kuznetsova & Markus, 2022 for individual compensatory trends in the long V2 shortening).

We can hypothesise that this sound change affects the abstract maximal limit for the overall duration of the foot rather than directly individual segments. Disyllables and shorter trisyllables still fit the new limit, and so their segmental quantities are preserved. Longer trisyllables do not fit it anymore, and so their quantity becomes affected in the position of V_2, the most sensitive to poly-subconstituent shortening in Finnic languages. Contrarily, disyllables and trisyllables with the shortest foot nucleus (i.e., the light feet) manifest prominent lengthening of V_2, which brings their overall duration closer to that of heavy feet and also works towards an isochronic pattern. Whether these processes of shortening and lengthening of unstressed V_2 are regulated by any teleological tendency toward foot isochrony or simply by foot stress, is a question. In any case, from the loss of V_2 contrast in Soikkola Ingrian trisyllables we can presume that compensatory phonetic effects can in principle overrun phonological length contrasts even in the languages with complex quantity, contrary to what has been claimed on the basis of Standard Finnish and Estonian.

7. Conclusion and outlook

Our study explored compensatory acoustic effects in 22 Soikkola Ingrian trisyllables, the four shortest of which were compared to respective disyllables. It confirmed earlier phonetic results on Soikkola Ingrian and other Finnic languages and added some novel findings concerning trisyllabic feet. Compensatory shortening effects in the first and second syllable of trisyllables are generally stronger than in the previous research conducted mostly on disyllables. This indicates a tendency for poly-subconstituent (both polysyllabic and polysegmental) shortening in trisyllables. Compression is not equally strong in different foot positions (as stated also by Eek, 1990, p. 253). While all compensatory effects in the first syllable are purely phonetic, the effect on the second syllable vowels is phonological: second syllable long vowels have shortened in all types of the trisyllabic heavy foot except the shortest one. On the other hand, the trisyllabic light foot, like the disyllabic light foot, manifests the prominent lengthening of its phonologically short second syllable vowel. The compensatory effects have overrun the phonological quantity in this position. As a result, the vocalic length contrast is now entirely lost from the second syllable of trisyllables and vowel duration is significantly inversely related to the first syllable length. In this, Soikkola Ingrian trisyllables have reached a new innovative typological stage of prosodic development comparable to that of other languages with ternary consonantal quantity: Estonian, Livonian, and Saami. Disyllables, being shorter feet, still maintain the length contrast of second syllable vowels. Our study confirms that the shortest Soikkola Ingrian disyllables manifest the cross-linguistically "purest" case of consonantal ternary quantity (i.e., the most independent of other durational effects).

We have discussed how the compensatory effects observed in our study might be especially strong in the chosen phrase-final position, due to a potential influence of boundary lengthening and stronger metrical prominence. For example, the lengthening of short V_2 in the light foot (the "half-long" vowel), which is most likely a non-segmental phenomenon related to stress, might be stronger here than in other prosodic positions. The phonological shortening of long V_2 in trisyllables, however, cannot be accounted for just by the phrasal position and the phonetic effects related to it. Whether this and other observed compensatory effects indicate a tendency toward foot isochrony, is an open issue, due to the controversy around the term itself and the presence of anticompensatory effects which disrupt the "ideal" isochronic patterns in Soikkola Ingrian (cf. also Kuznetsova, accepted),

Our results are potentially relevant for various phonological frameworks. Non-local prosodic patterns are not yet well-integrated in the models of Articulatory Phonology / Task Dynamics (Tilsen, 2019; Krivokapić, 2022), and those models are apparently not yet sufficiently specified to make concrete predictions in the case of complex durational patterns characteristic of Finnic languages (cf. Turk & Shattuck-Hufnagel, 2020, Chapter 6). Trisyllabic feet also present challenges for formal phonological accounts of metrical feet (cf. Martínez-Paricio & Kager, 2015, 2021; Torres-Tamarit & Jurgec, 2015; Golston, 2021). Ternary quantity contrasts are likewise challenging for formal phonological modelling (viz. Prehn, 2012; Kuznetsova, 2018b). Experimental evidence on ternary quantity contrasts and ternary feet is, therefore, of high theoretical importance.

Finnic quantity is one of the most complex quantity systems in the world, and Soikkola Ingrian has one of the most complex quantity systems in the Finnic space, if evaluated by a combination of length contrasts in phonology, quantity alternations in morphology, and structural compensatory and anticompensatory effects on duration and length. This study on Soikkola Ingrian brings us to a better understanding of such systems and the challenges they present to existing phonetic and phonological theories, raising our typological awareness of how quantity can work.

Author contributions

The design and the background of the study were developed by NK; the data were collected by NK, IB, and EM, segmented by IB with help of EM, extracted and processed by IB. IB conducted initial statistical analysis and data visualisation; NK was responsible for the selection of all final models and all final visualisation of data (apart from Appendix 3 performed by IB). The draft text was written by NK, revised and edited by NK with a help of EM. NK is the corresponding author and is responsible for the definitive version of the text.

Acknowledgements

We thank all Soikkola Ingrian speakers for working with us for many years. We are also extremely grateful to the editor-in-chief Taechong Cho, the associate editor Jeff Mielke, and several anonymous reviewers who have helped us to improve this paper significantly, and to Kurt Erbach and Sarah Bigi for proofreading.

Appendix 1

Quantity patterns in the nuclei of the 1–3-syllabic foot, attested in existing sources of the 19-20th cc.

Foot type	Type of C_2+V_2	Type of a sequence in the first syllable				
		(C)V-	(C)Vː-	(C)V_1V_2-	(C)VR-	(C)VːR-
Monosyllabic	—	no[1]	mā[2]	sai[3]	ken[4]	sūr[5]
	only C	nüt[6]	māt[7]	sait[8]	hänt[9]	sūrt[10]
Disyllabic	C + V	tapa [ˈtab̥aː][11]	lōta[12]	aita[13]	karta[14]	kārto[15]
	C + Vː		mātā[16]	aitās[17,a]	kartās[18,a]	*raentǟ*[b]
	(Cˑ + V)					
	Cˑ + Vː	tap̆pā[19]	lōt̆tā[20]	ait̆tā[21]	kart̆tā[22]	kārt̆tō[23]
	Cː + V	natta[24]	vōtta[25]	aitta[26]	kartta[27]	
	Cː + Vː	tappā[28]	nōttā[29]	aittā[30]	karttā[31]	
Trisyllabic	C + V	lakata [ˈłag̊aːd̥a][32]	vīkate[33]	leikata[34]	harkata[35]	*kērsim(m)ä*[c]
	(C + Vː)					
	Cˑ + V		vōt̆tava[36]	voit̆teli[37]	murk̆kina[38]	vā̃nt̆teli[39]
	Cˑ + Vː	mat̆tāla[40]	sūt̆tīm(m)a[41]	hoit̆tīm(m)a[42]	kerk̆kīm(m)ä[43]	vā̃nt̆tīm(m)ä[44]
	Cː + V	kattila[45]	ōttele[46]	voitteli[47]	markkoja[48]	
	Cː + Vː	kattīm(m)a[49]	mūttīm(m)a[50]	toittīm(m)a[51]	harkkām(m)a[52]	

Notes: [a]Only structures with late compensatory lengthening in the 2nd closed syllable are attested, e.g., **aitassa* > *aitās*. [b]Only examples with a "contracted" diphthong in the 1st syllable, previously divided by a syllabic boundary, are attested: *raentā* 'milk_pail:PRT' (varies in our data with *raent̆tǟ*), *tüärtǟ* 'daughter:PRT:3POS' (Sovijärvi, 1944, p. 40), *näöltǟ* 'face:PRT:3POS' (Laanest, 1986, p. 16), cf. Proto-Finnic **näköltäsen* 'face:PRT:3POS' > trisyllable *nä.öl.tǟ* > disyllable *näöl.tǟ*. [c]In the present phonetic study, bifoot words with a long 3rd syllable vowel (*vä̃ntelȫ* [ˈväːnd̥eˌlöː] 'twist:3SG') are used instead, but later we became aware also of true trisyllabic feet like *kērsim(m)ä* 'circle:PST:1PL' (Nirvi, 1971, p. 163) but with /s/ instead of a stop as C_2.

Glosses: [1]'but', [2]'land', [3]'get:PST:3SG', [4]'who', [5]'big', [6]'now', [7]'land:PL', [8]'get:PST:2SG', [9]'he:PRT', [10]'big:PRT', [11]'kill:IMP', [12]'broom', [13]'fence', [14]'baking_sheet', [15]'rainbow', [16]'sleep:IMPS', [17]'barn:IN', [18]'map:IN', [19]'catch:3SG', [20]'broom:PRT', [21]'fence:PRT', [22]'baking_sheet:PRT', [23]'rainbow:ILL', [24]'slime', [25]'year:PRT', [26]'barn', [27]'map', [28]'kill:3SG', [29]'seine', [30]'barn:PRT', [31]'map:PRT', [32]'sweep_floor:INF', [33]'scythe', [34]'cut:INF', [35]'step:INF', [36]'leak:PTC.PRS.ACT', [37]'smear:PST:3SG', [38]'breakfast', [39]'twist:PST:3SG', [40]'low', [41]'judge:1PL', [42]'beware:1PL', [43]'be_in_time:1PL', [44]'turn:1PL', [45]'cauldron', [46]'wait:IMP', [47]'struggle:PST:3SG', [48]'postage_stamp:PL:PRT', [49]'cover:1PL', [50]'change_oneself:1PL', [51]'feed_oneself:1PL', [52]'step:1PL'.

Appendix 2

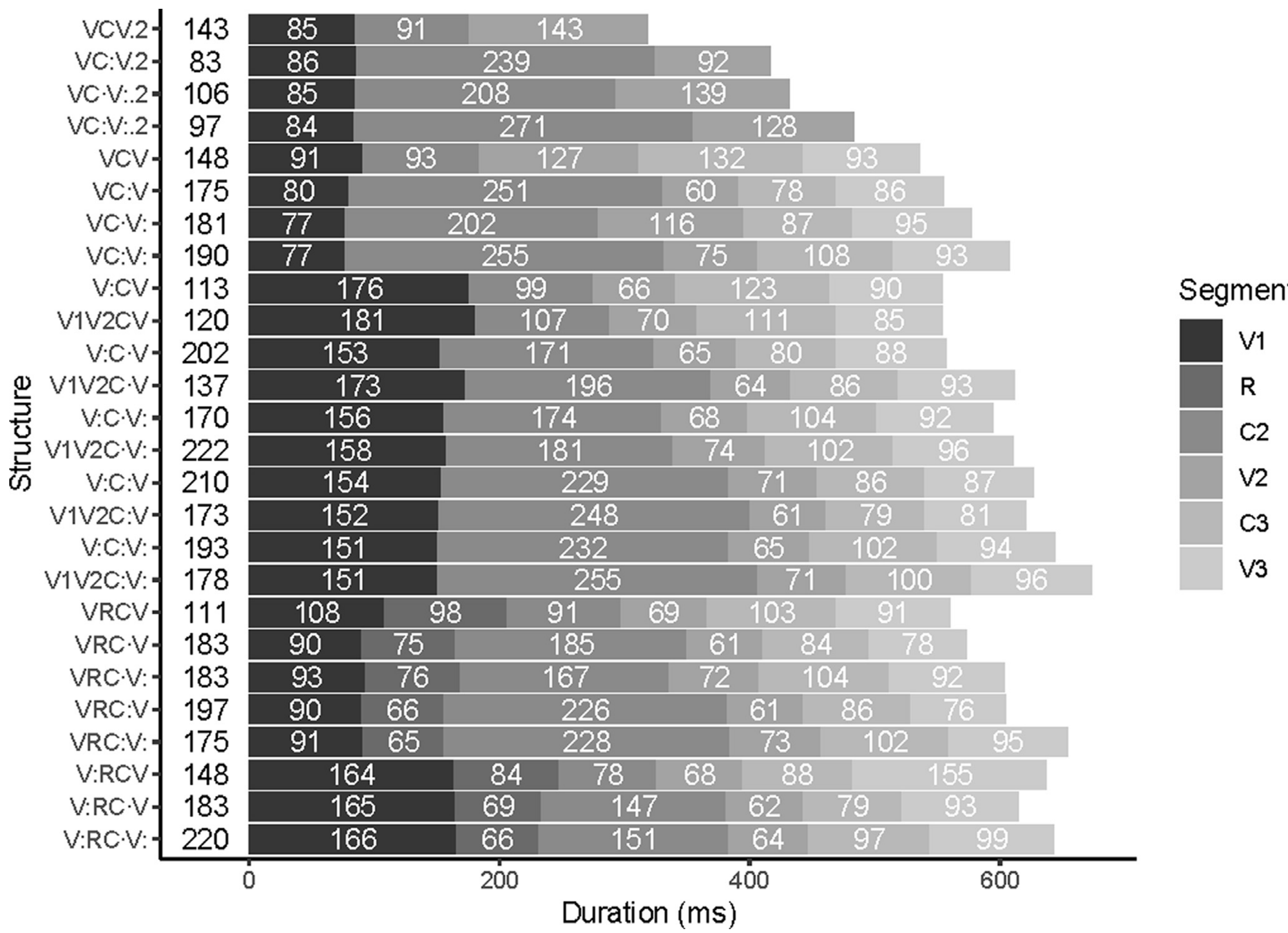
Sociolinguistic data on the speakers and the amount of collected data in Questionnaires 1–3 (labelled as “q1-q3”). Village names are given in Soikkola Ingrian together with their official Russian names in parentheses. Numbers of tokens placed in parentheses refer to the subset of STUDY 1 data used for STUDY 2.

Speaker code	Sub-dialect	Birth year	Birth place	Mother born	Father born	Recording place	STUDY 1 (n of tokens)			STUDY 2 (n of tokens)	
							Cˑ and Cː (q1)		C (q2)	3-syllables (q1-2 subset)	2-syllables (q3)
AL	north	1933	Hamm̆āla (Gamolovo)	Hamm̆āla (Gamolovo) (died early)	Ukraine	Voloitsa (Val’anicy)	267	342	210	(114)	165
AG	north	1936	Rep̆pōla (Repino)	no data	n./d.; died in 1937	Voloitsa (Val’anicy)	278	349	73	(103)	161
ST	north	1924	Mättüisī (Gorki)	Mättüisī (Gorki)	Mättüisī (Gorki)	Savimäki (Glinki)	298	239	52	(166)	103
LM	north	1930	Rep̆pōla (Repino)	Tarinaiššī (Andreevš’ina)	Rep̆pōla (Repino)	Mättüisī (Gorki)	331	356	90		
EI	transitory	1929	Vīstinä (Vistino)	Mättüisī (Gorki)	Vīstinä (Vistino)	Rūtsia (Ruč’ji)	317	395	215		
Tokens across STUDY 1 and 2 (q1-3): 4241						**Total:**	**1491**	**1681**	**640**	**(383)**	**429**

Notes: Questionnaire 1 (q1) contained structures nr. 36–52 from Appendix 1, and Questionnaire 2 contained structures nr. 32–35 and one bifoot structure *vä̅ntelȫ* ‘twist:3SG’ with a final long vowel, to cover the foot nucleus Vː RCV (see Ftn. *(c)* to Appendix 1). A consonant in the beginning of the third syllable (C_3) was usually a singleton. In cases where words with singletons were not available, we also included words with an original long geminate (**mm* or **ll*), which is often shortened up to a singleton in contemporary realisations (rendered in tokens in Supplementary materials as in *sūt̆tī**m(m)**a*). Each structure in Questionnaires 1–2 was included in five phrases with different carrier words. Each phrase was pronounced from 4 to 7 times by each of the five speakers (Questionnaire 2 contained less data than Questionnaire 1). STUDY 2 (Questionnaire 3) compared trisyllabic structures nr. 32, 40, 45, 49 from Appendix 1 (383 tokens) with disyllabic structures nr. 11, 19, 24, 28 (429 tokens) in three of the five speakers (AL, AG, ST). Each of the four disyllabic structures included 7–12 tokens pronounced by speakers from two to twelve times.

Appendix 3

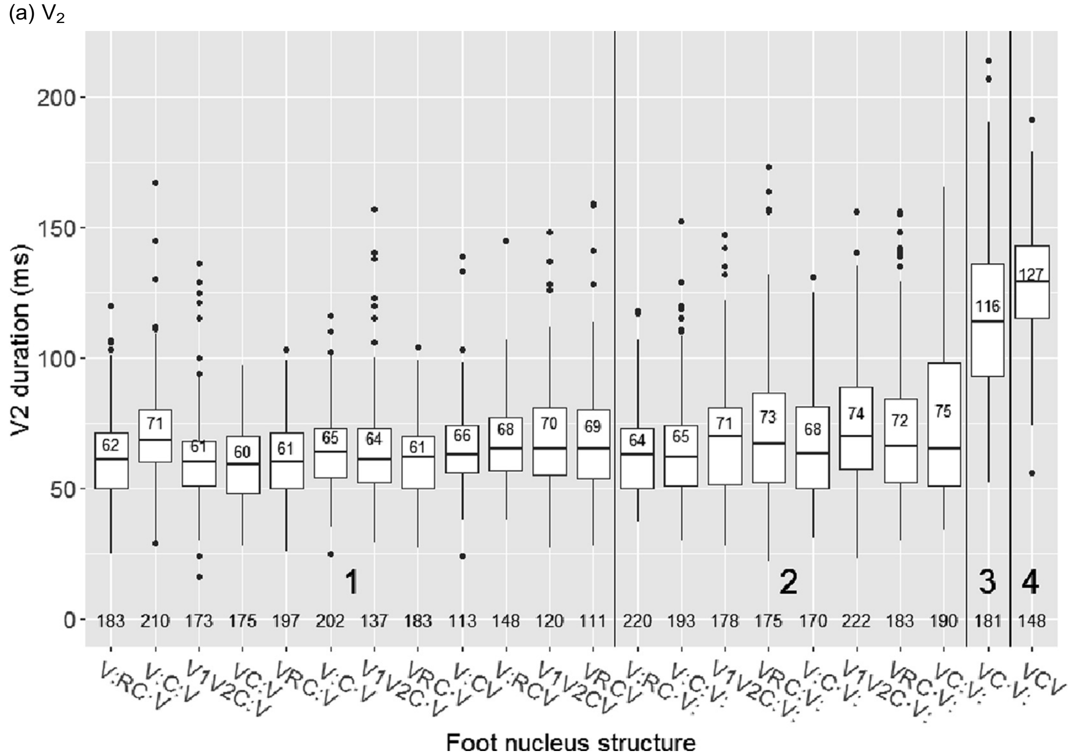
Raw mean durations of segments (apart from the word-initial consonant C_1) pooled over all tokens in all speakers in all structures of STUDY 1–2 and reported in ms in each box. Structures are coded by their foot nucleus types, and the number of tokens is given after the structure code before the respective bar. The first four are the disyllables (studied only in three speakers), followed by the 22 trisyllables (studied in five speakers).



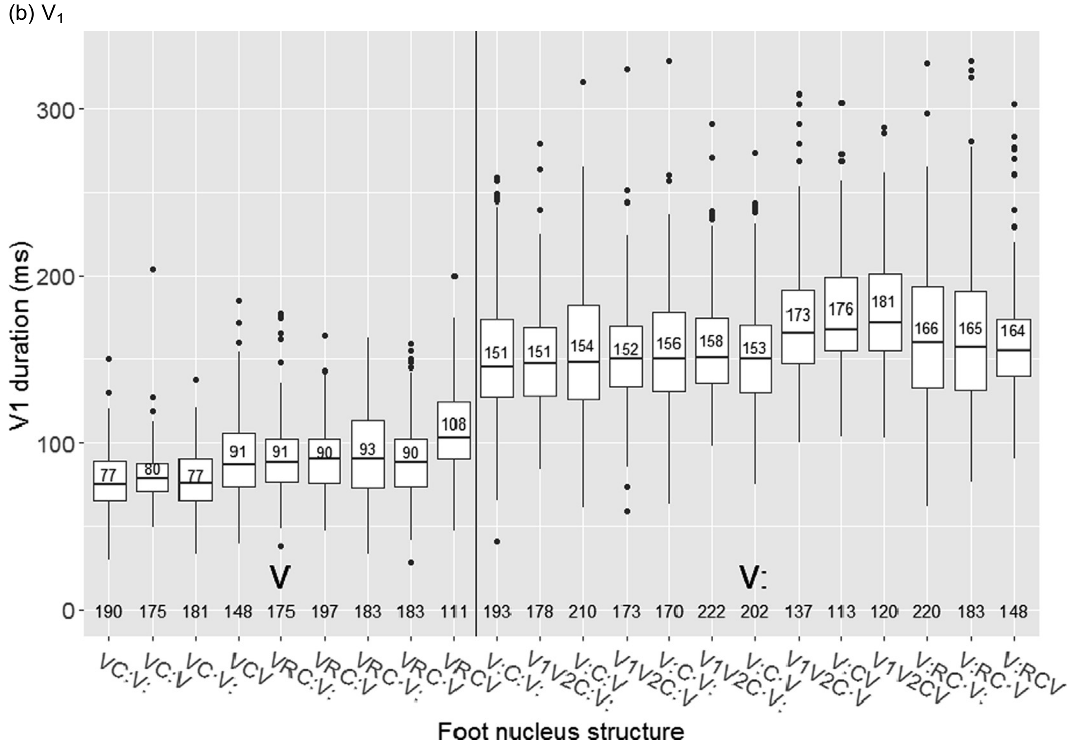
Appendix 4

Raw mean duration (in ms) of V_2, V_1, C_2, R in the 22 trisyllabic structures of STUDY 1. Boxes are coded by the foot nucleus types reported in Appendix 1. Raw means are reported within boxes, and the number of tokens is given below each box at the bottom. *(a)* Numbers from 1 to 4 code the four synchronic durational types of V_2 discussed in Section 5.2 *(b-c)* Structures are grouped by the phonological length of C_2 and V_1 respectively.

(a) V_2

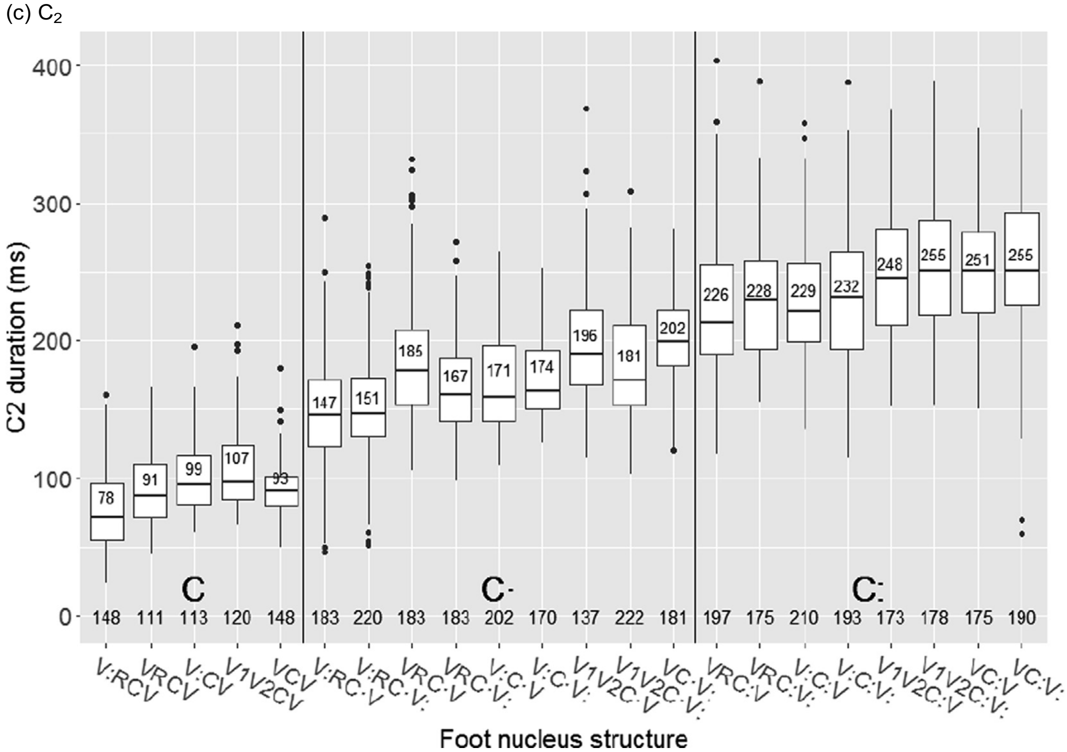


(b) V_1

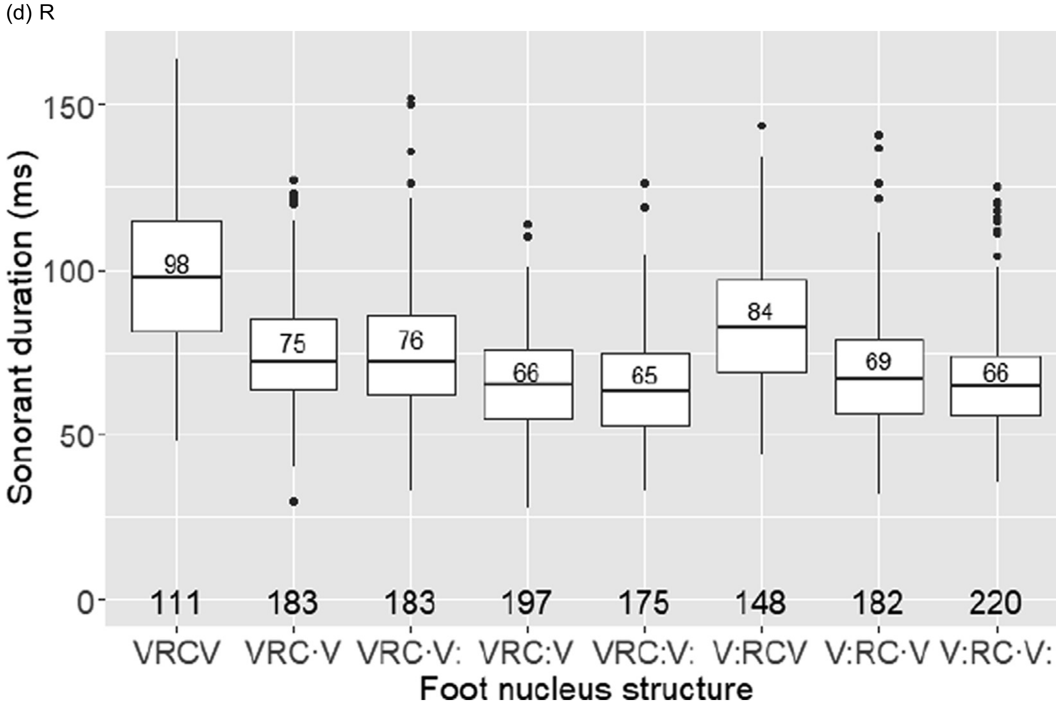


(continued on next page)

(c) C_2


C2 duration (ms)
0
100
200
300
400
78
91
99
107
93
147
151
185
167
171
174
196
181
202
226
228
229
232
248
255
251
255
C
C·
C:
148 111 113 120 148 183 220 183 183 202 170 137 222 181 197 175 210 193 173 178 175 190
V:RCV
VRCV
V:CV
V1V2CV
VCV
V:RC·V
V:RC·V:
VRC·V
VRC·V:
V:C·V
V:C·V:
V1V2C·V
V1V2C·V:
VC·V
VC·V:
VRC:V
VRC:V:
V:C:V
V:C:V:
V1V2C:V
V1V2C:V:
VC:V
VC:V:
Foot nucleus structure

(d) R


Sonorant duration (ms)
0
50
100
150
98
75
76
66
65
84
69
66
111 183 183 197 175 148 182 220
VRCV
VRC·V
VRC·V:
VRC:V
VRC:V:
V:RCV
V:RC·V
V:RC·V:
Foot nucleus structure

Appendix 5

Summary of modelled interaction between the foot nucleus type and the number of syllables in the foot for V_2 and C_2 in STUDY 2. Satterthwaite's approximation of the degrees of freedom and a confidence level 95% (LCL and UCL refer to the lower and upper confidence level respectively) are used for *t*- and *p*-statistics. *Part I* reports the least squares means in V_2 (upper part) and C_2 (lower part) within various foot structures. Coding such as “2.VCV” or “3.VCV” refers to the number of syllables in the foot and the structure of the foot nucleus. *Part II* reports pairwise differences between the least squares means in V_2 (left part) and C_2 (right part) in various pairs of foot structures. Comparisons from top to bottom: pairs of disyllabic feet with different types of foot nucleus; pairs of trisyllabic feet with different types of foot nucleus; pairs of disyllables vs. trisyllables with identical types of foot nucleus.

I. Least squares means of V2 and C2 (in ms)							**II. Pairwise differences of least squares means (in ms)**											
									V2					C2	C2	C2	C2	C2
Position	Structure	Mean	SE	df	LCL	UCL	Structure	Contrast	Mean	SE	df	t-value	p-value (>|t|)	Estimate	SE	df	t-ratio	p-value (>|t|)
V2	2.VCV	143.6	12.3	2.37	97.9	189.2	2-syllabic	VCV-VC·V:	4.6	6.03	45.2	0.8	0.4461	-119.1	7.7	46.6	-15.47	<.0001
	2.VC·V:	138.9	12.7	2.68	95.8	182.1		VC·V:-VC:V:	8.1	6.59	46.0	1.2	0.2241	-66.0	8.4	47.7	-7.851	<.0001
	2.VC:V:	130.8	12.6	2.60	87.0	174.6		VCV-VC:V:	12.8	5.80	53.4	2.2	0.0321	-185.1	7.3	52.6	-25.4	<.0001
	2.VC:V	90.4	12.6	2.66	47.1	133.7		VCV-VC:V	53.2	5.95	53.1	8.9	<.0001	-150.2	7.5	51.3	-20.04	<.0001
	3.VCV	127.7	12.7	2.73	84.8	170.5		VC·V:-VC:V	48.6	6.73	46.3	7.2	<.0001	-31.1	8.6	47.2	-3.622	0.0007
	3.VC·V:	106.5	12.7	2.67	63.3	149.8		VC:V:-VC:V	40.4	6.53	53.0	6.2	<.0001	34.9	8.2	52.0	4.239	0.0001
	3.VC:V:	66.4	12.6	2.65	23.1	109.8	3-syllabic	VCV-VC·V:	21.2	6.89	49.1	3.1	0.0035	-97.5	8.7	51.9	-11.2	<.0001
	3.VC:V	53.0	12.7	2.75	10.2	95.7		VC·V:-VC:V:	40.1	6.72	43.8	6.0	<.0001	-50.3	8.6	47.7	-5.858	<.0001
C2	2.VCV	85.8	13.1	2.56	39.8	132.0		VCV-VC:V:	61.3	6.84	47.3	9.0	<.0001	-147.8	8.7	50.0	-17.02	<.0001
	2.VC·V:	204.9	13.7	3.10	162.0	248.0		VCV-VC:V	74.7	7.04	48.0	10.6	<.0001	-149.4	8.9	50.3	-16.73	<.0001
	2.VC:V:	270.9	13.5	2.90	227.1	315.0		VC·V:-VC:V	53.6	6.92	44.5	7.7	<.0001	-51.8	8.8	48.1	-5.867	<.0001
	2.VC:V	236.0	13.6	3.01	192.7	279.0		VC:V:-VC:V	13.5	6.86	42.7	2.0	0.0563	-1.5	8.8	46.2	-0.174	0.8627
	3.VCV	89.2	13.7	3.11	46.4	132.0	VCV	2 vs. 3 syllables in the foot	15.9	6.17	54.0	2.6	0.0129	-3.5	7.7	54.0	-0.4	0.6559
	3.VC·V:	186.8	13.7	3.05	143.7	230.0	VC·V:		32.4	6.78	43.0	4.8	<.0001	18.1	8.7	46.5	2.1	0.0425
	3.VC:V:	237.1	13.7	3.04	193.9	280.0	VC:V:		64.4	6.54	47.1	9.8	<.0001	33.8	8.3	49.3	4.1	0.0002
	3.VC:V	238.6	13.8	3.18	196.0	281.0	VC:V		37.4	6.89	48.1	5.4	<.0001	-2.6	8.8	49.0	-0.3	0.7677

Appendix 6

Summary of modelled length contrasts in the V_1, C_2, and V_2 of trisyllables (STUDY 1). Least squares means (*Part I)* and pairwise differences between them (*Part II)* are reported. Satterthwaite's approximation of the degrees of freedom and a confidence level 95% (LCL and UCL refer to the lower and upper confidence level respectively) are used for *t*- and *p*-statistics.

I. Least squares means of length classes (in ms)							**II. Pairwise differences of least squares means (in ms)**							
Position	Length	Mean	SE	df	LCL	UCL	Contrast	Estimate	SE	df	LCL	UCL	t-ratio	p-value(>|t|)
V1	V1	93.9	9.54	6.70	71.1	116.6	**V1-V1**ː	-86.6	4.85	93.0	-96.3	-77.0	-17.9	<2.2e-16
	V1ː	180.5	9.68	7.10	157.7	203.3								
C2	C2	94.0	13.30	5.00	59.8	128.2	**C-C**ˑ	-82.8	4.52	119.7	-91.8	-78.1	-18.3	<2.2e-16
	C2ˑ	176.8	12.94	4.49	142.4	211.3	**C-C**ː	-140.2	4.58	121.5	-149.3	-134.3	-30.6	<2.2e-16
	C2ː	234.2	12.94	4.49	199.8	268.7	**C**ˑ**-C**ː	-57.4	3.43	119.1	-64.2	-51.0	-16.7	<2.2e-16
***V2** (original phonological length: Types 1&4 vs. 2&3)	*V2	73.3	4.46	6.40	62.6	84.1	***V2-*V2**ː	-8.5	4.27	6.3	-18.9	1.8	-2.0	0.09004
	*V2ː	81.9	7.40	5.20	63.1	100.7								

Appendix 7

Detailed statistics on modelled effects from STUDY 1 reported in Section 5.4 (with a reference to corresponding figure numbers). Differences between least squares means are given. Kenward-Roger's approximation of the degrees of freedom and a confidence level 95% are used for *t*- and *p*-statistics.

Fig.	**Variable**	**Factor 1 value**	**Contrast in Factor 2 values**	**Estimate**	**SE**	**df**	**t-ratio**	**p-value**
7a	V1 duration	V1 = short	C2-C2ˑ	16.4	4.64	129	3.5	0.0006
			C2-C2ː	12.0	4.41	125	2.7	0.0072
			C2ˑ-C2ː	-4.4	3.43	119	-1.3	0.2048
		V1 = long	C2-C2ˑ	18.1	3.64	116	5.0	<.0001
			C2-C2ː	27.8	3.96	111	7.3	<.0001
			C2ˑ-C2ː	10.7	3.28	117	3.3	0.0015
7b	C2 duration	C2 = C	V1-V1ː	3.2	5.97	116	0.5	0.5922
		C2 = Cˑ	V1-V1ː	25.1	4.01	112	6.2	<.0001
		C2 = Cː	V1-V1ː	20.6	5.13	107	4.0	0.0001
8	C2 duration	C2 = C	no-yes R	19.0	5.90	117	3.2	0.0017
		C2 = Cˑ	no-yes R	29.4	4.19	111	7.0	<.0001
		C2 = Cː	no-yes R	34.0	4.95	117	6.9	<.0001
9a	R duration		V1-V1ː	11.3	2.1	44.5	5.3	<.0001
9b	R duration		C2-C2ˑ	19.2	2.2	42.9	8.7	<.0001
			C2-C2ː	29.7	2.7	45.9	11.2	<.0001
			C2ˑ-C2ː	10.6	2.2	45.1	4.9	<.0001

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2023.101246.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003.

Asu-Garcia, E. L., Lippus, P., Pajusalu, K., & Teras, P. (2016). Eesti keele hääldus [Pronunciation of the Estonian language]. *Tartu Ülikool Kirjastus*. https://dspace.ut.ee/handle/10062/57960.

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. https://doi.org/10.1016/j.jml.2007.12.005.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01.

Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer* (5.3.81). www.praat.org.

Eek, A. (1974). Observations on the duration of some word structures: I. In A. Eek (Ed.), *Estonian papers in phonetics (EPP) 1974* (pp. 18–32). Academy of sciences of the Estonian SSR; Institute of language and literature.

Eek, A. (1975). Observations on the duration of some word structures: II. In A. Eek (Ed.), *Estonian papers in phonetics (EPP) 1975* (pp. 7–55). Academy of sciences of the Estonian SSR; Institute of language and literature.

Eek, A. (1990). Units of temporal organisation and word accents in Estonian. *Linguistica Uralica, 26*(4), 251–263.

Eek, A., & Meister, E. (2003). Foneetilisi katseid ja arutlusi kvantiteedi alalt (I). Häälikukestusi muutvad kontekstid ja välde. [Phonetic experiments and discussions from the field of quantity (I). The contexts which change segmental durations and the degree of quantity]. Keel ja kirjandus, 11, 12; 815–837; 904–918. https://www.digar.ee/arhiiv/et/download/143198.

Eek, A., & Meister, E. (2004). Foneetilisi katseid ja arutlusi kvantiteedi alalt (II). Takt, silp ja välde [Phonetic experiments and discussions from the field of quantity (II). Rhythmic group, syllable, and quantity]. Keel ja kirjandus, 4, 5; 251–271, 337–357. https://www.digar.ee/arhiiv/en/download/143202.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage.

Genetz, A. (1881). *Suomen kielen äänne- ja muoto-oppi ynnä runousoppi: Oppikouluja varten [Phonetics and morphology, as well as poetry study, of the Finnish language: For grammar schools]*. Holm: K. E.

Golston, C. (2021). No stress system requires recursive feet. *Catalan Journal of Linguistics, 20*, 9–35. https://doi.org/10.5565/rev/catjl.333.

Gordon, M. K. (2009). Prominence and gemination in Ingrian. *Linguistica Uralica, 45*(2), 81–100. https://doi.org/10.3176/lu.2009.2.01.

Hakulinen, L. (1922). Pari suomen murteiden vokaalinkestoseikkaa [A couple of Finnish dialectal vowel duration phenomena]. *Virittäjä, 26*, 49–59.

van der Hulst, H. (2010). Representing accent. *Phonological Studies, 13*, 117–128.

van der Hulst, H. (2014). Representing rhythm. In H. van der Hulst (Ed.), *Word stress: Theoretical and typological issues* (pp. 325–365). Cambridge University Press. https://doi.org/10.1017/CBO9781139600408.015.

Iivonen, A., Sovijärvi, A., & Aulanko, R. (1990). *Foneettisen kirjoituksen kehitys ja nykytila: Kansainvälinen foneettinen aakkosto (IPA)*. Suomalais-ugrilainen tarkekirjoitus (SUT) [Development and current state of phonetic transcription: International Phonetic Alphabet (IPA). Finno-Ugric phonetic transcription (FUT)]: Helsingin yliopisto.

Krivokapić, J. (2022). Prosody in articulatory phonology. In J. A. Barnes & S. Shattuck-Hufnagel (Eds.), *Prosodic theory and practice* (pp. 213–246). The MIT Press. https://doi.org/10.7551/mitpress/10413.003.0008.

Krull, D. (1997). Prepausal lengthening in Estonian: Evidence from conversational speech. In I. Lehiste & J. Ross (Eds.), *Estonian prosody: Papers from a symposium: International Symposium on Estonian Prosody, Tallinn, Estonia, October 29–30, 1996: Proceedings* (pp. 136–148). Institute of Estonian language.

Krull, D. (1999). Foot isochrony in Estonian. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS-14), San Francisco, CA, US, August 1-7* (pp. 1063–1066). University of California.

Kuznecova, N. V. (2009a). Fonologičeskije sistemy ižorskih dialektov [Phonological systems of Ingrian dialects]. *Diss.kand. filol. nauk [Candidate thesis]]. Institut lingvističeskih issledovanij RAN [Institute for Linguistic Studies, Russian Academy of Sciences]*.

Kuznecova, N. V. (2009b). Suprasegmentnaja fonologija sojkinskogo dialekta ižorskogo jazyka v tipologičeskom aspekte [Suprasegmental phonology of the Soikkola dialect of Ingrian in the typological aspect]. *Voprosy Jazykoznanija, 5*, 18–47 https://vja.ruslang.ru/en/archive/2009-5/18-47.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Kuznetsova, N. (2015). Two phonological rarities in Ingrian dialects. In M. Hilpert, J.-O. Östman, C. Mertzlufft, M. Riessler, & J. Duke (Eds.), *New trends in Nordic and general linguistics* (pp. 91–117). De Gruyter. Doi: 10.1515/9783110346978.91.

Kuznetsova, N. (2016). Evolution of the non-initial vocalic length contrast across the Finnic varieties of Ingria and adjacent areas. *Linguistica Uralica, 52*(1), 1–25. https://doi.org/10.3176/lu.2016.1.01.

Kuznetsova, N. (2018a). What Danish and Estonian can show to a modern word-prosodic typology. In R. Goedemans, J. Heinz, & H. van der Hulst (Eds.), *The study of word stress and accent: Theories, methods and data* (pp. 102–143). Cambridge University Press. 10.1017/9781316683101.005.

Kuznetsova, N. (2018b). Estonian word prosody on the Procrustean bed of morae. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric. Linguistics, 9*(1), 209–244. https://doi.org/10.12697/jeful.2018.9.1.09.

Kuznetsova, N. (accepted). Typology, chronology, and phonetic mechanisms of Finnic secondary gemination in the light of Soikkola Ingrian acoustic data. In O. Niebuhr & M. Svensson Lundmark (Eds.), *Proceedings of the 13th International Conference of Nordic Prosody: Phonetics, phonology, and technology for and of Scandinavian languages*. Sciendo. https://sciendo.com/book/9788366675728.

Kuznetsova, N., Brodskaya, I., & Markus, E. (2023). Interaction of quantity, foot structure, and stress in the 2nd/3rd syllable sonorants of Soikkola Ingrian trisyllables. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, to appear.

Kuznetsova, N., & Markus, E. (2022). Ongoing vowel shortening in vanishing Soikkola Ingrian: Challenges for description, codification, and typology. In S. Frota & M. Vigário (Eds.), Proceedings of the 11th International Conference on Speech Prosody, 23-26 May, 2022 (pp. 327–331). Universidade de Lisboa. https://doi.org/10.21437/SpeechProsody.2022-67.

Kuznetsova, N., Markus, E., & Muslimov, M. (2015). Finnic minorities of Ingria: The current sociolinguistic situation and its background. In H. F. Marten, M. Rießler, J. Saarikivi, & R. Toivanen (Eds.), *Cultural and linguistic minorities in the Russian Federation and the European Union: Comparative studies on equality and diversity* (pp. 127–167). Springer. https://doi.org/10.1007/978-3-319-10455-3_6.

Kuznetsova, N., & Verkhodanova, V. (2019). Phonetic realisation and phonemic categorisation of the final reduced corner vowels in the Finnic languages of Ingria. *Phonetica, 76*(2–3), 201–233. https://doi.org/10.1159/000494927.

Laanest, A. (1986). *Isuri keele ajalooline foneetika ja morfoloogia [Historical phonetics and morphology of the Ingrian language]*. Keele ja Kirjanduse Instituut: Valgus; ENSV Teaduste Akadeemia.

Laanest, A. (1987). Zur Phonologie der ingrischen [On the phonology of Ingrian]. In *Studien zur Phonologie und Morphonologie der uralischen Sprachen: Eisenstadt, 28. Juni—1. Juli 1984: Akten der dritten Tagung für uralische Phonologie* (pp. 285–293). Verband der wissenschaftlichen Gesellschaften Österreichs.

Laanest, A. (1993). Ižorskij jazyk [The Ingrian language]. In J. S. Eliseev, K. E. Majtinskaja, & O. I. Romanova (Eds.), *Jazyki mira: Ural'skije jazyki [Languages of the world: Uralic languages]* (pp. 55–63). Nauka.

Lahiri, A., & Fikkert, P. (1999). Trisyllabic shortening in English: Past and present. *English Language and Linguistics, 3*(2), 229–267. https://doi.org/10.1017/S1360674399000234.

Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. *American Studies in Uralic Linguistics, 1*, 21–82.

Lehiste, I. (1968). Vowel quantity in word and utterance in Estonian. In M. Kahla, P. Ravila, & A. Räisänen (Eds.), *Congressus Secundus Internationalis Fenno-Ugristarum, Helsingiae habitus 23.-28.VIII.1965. Pars 1: Acta linguistica* (pp. 293–303). Societas Fenno-Ugrica.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *The Journal of the Acoustical Society of America, 51*(6B), 2018–2024. https://doi.org/10.1121/1.1913062.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5*(3), 253–263. https://doi.org/10.1016/S0095-4470(19)31139-8.

Lehiste, I. (1997). The structure of trisyllabic words. In I. Lehiste & J. Ross (Eds.), *Estonian prosody: Papers from a symposium. Proceedings of the International Symposium on Estonian Prosody, Tallinn, Estonia, October 29-30 1996* (pp. 149–164). Institute of Estonian language. https://doi.org/10.7551/mitpress/10413.003.0008.

Lehiste, I., Teras, P., Ernštreits, V., Lippus, P., Pajusalu, K., Tuisk, T., & Viitso, T.-R. (2008). *Livonian prosody*. Suomalais-ugrilainen seura.

Lehtonen, J. (1970). *Aspects of quantity in standard Finnish*. University of Jyväskylä.

Lehtonen, J., & Leskinen, H. (1973). Kvantiteettisuhteista läntisissä kaakkoismurteissa [On quantitative relations in western South-Eastern dialects]. *Virittäjä, 77*(4), 315–326 https://journal.fi/virittaja/article/view/36348.

Length, R. (2020). *Emmeans: Estimated marginal means, aka least-squares means* (R package version 1.4.6). https://CRAN.R-project.org/package=emmeans.

Leskinen, H. (1978). Kaksi kaakkoismurteiden ja inkeroismurteiden kestoseikkaa [Two durational facts of South-Eastern and Ingrian dialects]. *Virittäjä, 82*(2), 122–130 https://journal.fi/virittaja/article/view/36955.

Liiv, G. (1961). Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid [Duration and melodical types of Estonian vowels in three degrees of quantity]. *Keel Ja Kirjandus*. 7, 8; 412–424; 480–490.

Lippus, P. (2011). *The acoustic features and perception of the Estonian quantity system* [Doctoral dissertation, University of Tartu]. https://dspace.ut.ee/handle/10062/18923.

Lippus, P., Asu, E. L., Teras, P., & Tuisk, T. (2013). Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics, 41*(1), 17–28. https://doi.org/10.1016/j.wocn.2012.09.005.

Lippus, P., Pajusalu, K., & Teras, P. (2006). The temporal structure of penta- and hexasyllabic words in Estonian. In R. Hoffmann & H. Mixdorff (Eds.), Proceedings of the 3rd International Conference of Speech Prosody (pp. 759–762). TUDpress. https://www.isca-speech.org/archive/pdfs/speechprosody_2006/lippus06_speechprosody.pdf.

Loukina, A., Kochanski, G., Rosner, B., Keane, E., & Shih, C. (2011). Rhythm measures and dimensions of durational variation in speech. *The Journal of the Acoustical Society of America, 129*(5), 3258–3270. https://doi.org/10.1121/1.3559709.

Maddieson, I. (2018). Is phonological typology possible without (universal) categories? In L. M. Hyman & F. Plank (Eds.), *Phonological typology* (pp. 107–125). De Gruyter Mouton. https://doi.org/10.1515/9783110451931.

Markus, E. (2010). Primary and secondary geminates in Ingrian. *Linguistica Uralica, 46*(1), 38–52. https://doi.org/10.3176/lu.2010.1.03.

Markus, E. (2011). The phonetics and phonology of a disyllabic foot in Soikkola Ingrian. *Linguistica Uralica, 47*(2), 103–119. https://doi.org/10.3176/lu.2011.2.03.

Markus, E., Lippus, P., Pajusalu, K., & Teras, P. (2013). Three-way opposition of consonant quantity in Finnic and Saamic languages. In E. L. Asu-Garcia & P. Lippus (Eds.), *Nordic prosody: Proceedings of the XIth Conference, Tartu 2012* (pp. 225–234). Peter Lang Edition. https://doi.org/10.3726/978-3-653-03047-1.

Martínez-Paricio, V., & Kager, R. (2015). The binary-to-ternary rhythmic continuum in stress typology: Layered feet and non-intervention constraints. *Phonology, 32*(3), 459–504. https://doi.org/10.1017/S0952675715000287.

Martínez-Paricio, V., & Kager, R. (2021). In favour of layered feet: A response to Golston. *Catalan Journal of Linguistics, 20*, 37–55. https://doi.org/10.5565/rev/catjl.352.

McRobbie-Utasi, Z. (2007). The instability of systems with ternary length distinctions: The Skolt Saami evidence. In I. Toivonen & D. C. Nelson (Eds.), *Saami linguistics* (pp. 167–205). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.288.07mcr.

Nahkola, K. (1987). *Yleisgeminaatio: Äänteenmuutoksen synty ja vaiheet kielisysteemissä erityisesti Tampereen seudun hämäläismurteiden kannalta [Common gemination: the birth and phases of the sound change in the language system especially regarding the Häme dialects of the vicinity of Tampere]*. Suomalaisen kirjallisuuden seura.

Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics, 37*(1), 29–45. https://doi.org/10.1016/j.wocn.2008.08.002.

Nakai, S., Turk, A. E., Suomi, K., Granlund, S., Ylitalo, R., & Kunnari, S. (2012). Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics, 40*(6), 796–807. https://doi.org/10.1016/j.wocn.2012.08.003.

Nirvi, R. E. (1971). *Inkeroismurteiden sanakirja [The dictionary of Ingrian dialects]*. Suomalais-Ugrilainen Seura.

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences, 13*(1), 14–19. https://doi.org/10.1016/j.tics.2008.09.005.

Pajusalu, K., Lippus, P., Niit, E., Teras, P., & Viitso, T.-R. (2005). On the temporal structure of Estonian secondary-stressed feet. *Linguistica Uralica, 41*(2), 98–106. https://doi.org/10.3176/lu.2005.2.03.

Palander, M. (1987). *Suomen itämurteiden erikoisgeminaatio [Special gemination of Eastern Finnish dialects]*. Suomalaisen kirjallisuuden seura.

Pike, K. L. (1945). *The intonation of American English* (Vol. 1). University of Michigan Press.

Plüschke, M. (2013). *Peak alignment in Estonian [Doctoral dissertation]*. Ludwig-Maximilians-Universität München https://edoc.ub.uni-muenchen.de/15874/1/Plueschke_Mareike.pdf.

Porkka, V. (1885). *Ueber den Ingrischen Dialekt: Mit Berücksichtigung der übrigen finnisch-ingermanländischen Dialekte [On the Ingrian dialect: With consideration of other Finnic dialects of Ingria]*. J. C: Frenckell & Sohn.

Prehn, M. (2012). Vowel quantity and the fortis-lenis distinction in North Low Saxon. *LOT* https://dare.uva.nl/search?identifier=913ec558-f1af-42c4-9380-078dbfbb6efc.

R Core Team. (2020). *R: A language and environment for statistical computing* (4.0.0). R foundation for statistical computing. http://www.R-project.org/.

Rathcke, T. V., & Smith, R. H. (2015). Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies. *The Journal of the Acoustical Society of America, 137*(5), 2834–2845. https://doi.org/10.1121/1.4919322.

Remijsen, B., Ayoker, O. G., & Jørgensen, S. (2019). Ternary vowel length in Shilluk. *Phonology, 36*(1), 91–125. https://doi.org/10.1017/S0952675719000058.

Remijsen, B., & Gilley, L. (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *Journal of Phonetics, 36*(2), 318–344. https://doi.org/10.1016/j.wocn.2007.09.002.

Romano, A. (2020). An account on vowel reduction and deletion in Apulian and Lucanian dialects with reference to speech rhythm. *Italian Journal of Linguistics, 32* (1), 85–102. https://doi.org/10.26346/1120-2726-149.

Sadeniemi, M. (1949). Metriikkamme perusteet ja sovellutusta moderneihin ja antiikin mittoihin [The foundations of our metrics and application to modern and ancient measures]. *Suomalaisen Kirjallisuuden Seura.*.

Sammallahti, P. (1998). *The Saami languages: An introduction*. Davvi Girji.

Sovijärvi, A. (1944). *Foneettis-äännehistoriallinen tutkimus Soikkolan inkeroismurteesta [Phonetic and historical-phonological study on the Soikkola Ingrian dialect]*. Suomalaisen kirjallisuuden seura.

Suomi, K. (2009). Durational elasticity for accentual purposes in Northern Finnish. *Journal of Phonetics, 37*(4), 397–416. https://doi.org/10.1016/j.wocn.2009.07.003.

Suomi, K., Meister, E., Ylitalo, R., & Meister, L. (2013). Durational patterns in Northern Estonian and Northern Finnish. *Journal of Phonetics, 41*(1), 1–16. https://doi.org/10.1016/j.wocn.2012.09.001.

Suomi, K., Toivanen, J., & Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics, 31*(1), 113–138. https://doi.org/10.1016/S0095-4470(02)00074-8.

Suomi, K., Toivanen, J., & Ylitalo, R. (2008). Finnish sound structure: Phonetics, phonology, phonotactics and prosody. *University of Oulu* http://jultika.oulu.fi/files/isbn9789514289842.pdf.

Suomi, K., & Ylitalo, R. (2004). On durational correlates of word stress in Finnish. *Journal of Phonetics, 32*(1), 35–63. https://doi.org/10.1016/S0095-4470(03)00005-6.

Tilsen, S. (2019). Motoric mechanisms for the emergence of non-local phonological patterns. *Frontiers in Psychology, 10*, 2143. https://doi.org/10.3389/fpsyg.2019.02143.

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America, 134*(1), 628–639. https://doi.org/10.1121/1.4807565.

Torres-Tamarit, F., & Jurgec, P. (2015). Lapsed derivations: Ternary stress in harmonic serialism. *Linguistic Inquiry, 46*(2), 376–387. https://doi.org/10.1162/LING_a_00186.

Tuisk, T. (2015). The temporal structure of Livonian tri-, tetra- and pentasyllabic words. *Linguistica Uralica, 61*(3), 177. https://doi.org/10.3176/lu.2015.3.02.

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. https://doi.org/10.1515/lp-2013-0005.

Turk, A., & Shattuck-Hufnagel, S. (2020). Speech timing: Implications for theories of phonology, speech production, and speech motor control. *Oxford University Press*. https://doi.org/10.1093/oso/9780198795421.001.0001.

Türk, H., Lippus, P., Pajusalu, K., & Teras, P. (2019). The acoustic correlates of quantity in Inari Saami. *Journal of Phonetics, 72*, 35–51. https://doi.org/10.1016/j.wocn.2018.11.001.

Türk, H., Lippus, P., Pajusalu, K., & Teras, P. (2020). Temporal patterns of geminates in Inari Saami trisyllabic words. In *Proceedings of the 10th International Conference on Speech Prosody* (pp. 81–85). https://doi.org/10.21437/SpeechProsody.2020-17.

White, L. (2002). *English speech timing: A domain and locus approach* [Doctoral dissertation, University of Edinburgh]. http://www.cstr.ed.ac.uk/projects/eustace/fulltext/lswdiss.pdf.

White, L. (2014). Communicative function and prosodic form in speech timing. *Speech Communication, 63–64*, 38–54. https://doi.org/10.1016/j.specom.2014.04.003.

White, L., & Turk, A. E. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics, 38*(3), 459–471. https://doi.org/10.1016/j.wocn.2010.05.002.

Wickham, H. (2016). **ggplot2**: Elegant graphics for data analysis. *Springer International Publishing*. https://doi.org/10.1007/978-3-319-24277-4.

Wiik, K., & Lehiste, I. (1968). Vowel quantity in Finnish dissyllabic words. In M. Kahla, P. Ravila, & A. Räisänen (Eds.), *Congressus Secundus Internationalis Fenno-Ugristarum Helsingiae habitus 23.–28.VIII.1965. Pars I: Acta linguistica* (pp. 569–574). Societas Fenno-Ugrica.