

JOURNAL ARTICLE

# Context-dependent articulation of consonant gemination in Estonian

Helen Türk<sup>1,2</sup>, Pärtel Lippus<sup>1,2</sup> and Juraj Šimko<sup>2</sup>

<sup>1</sup> Institute of Estonian and General Linguistics, University of Tartu, Estonia, EE

<sup>2</sup> Institute of Behavioural Sciences, University of Helsinki, FI

Corresponding author: Helen Türk ([helen.turk@ut.ee](mailto:helen.turk@ut.ee))

The three-way quantity system is a well-known phonological feature of Estonian. In a number of studies it has been shown that quantity is realized in a disyllabic foot by the stressed-to-unstressed syllable rhyme duration ratio and also by pitch movement as the secondary cue. The stressed syllable rhyme duration is achieved by combining the length of the vowel and the coda consonant, which enables minimal septets of CVCV-sequences based on segmental duration. In this study we analyze articulatory (EMA) recordings from four native Estonian speakers producing all possible quantity combinations of intervocalic bilabial stops in two vocalic contexts (/a-i/ vs. /i-a/). The analysis shows that kinematic characteristics (gesture duration, spatial extent, and peak velocity) are primarily affected by quantity on the segmental level: Phonologically longer segments are produced by longer and larger lip closing gestures and, in reverse, shorter and smaller lip opening movements. Tongue transition gesture is consistently lengthened and slowed down by increasing consonant quantity. In general, both kinematic characteristics and inter-gestural coordination are influenced by non-linear interactions between segmental quantity levels as well as vocalic context.

**Keywords:** phonological quantity; three-way quantity; Estonian; articulation

## 1. Introduction

In the present paper we investigate the articulatory characteristics of the realization of the phonological quantity contrast in Estonian. We analyze the influence of quantity on the kinematic properties—duration, spatial extent, peak velocity—of the lip closing and opening gestures of a bilabial consonant and on the inter-vocalic transition gesture considering the impact of the articulatory nature of a word-initial consonant as well as the vocalic context on the kinematics of the gestures. In terms of quantity and context sensitivity, we also study the temporal coordination between the bilabial lip movement and the coproduced inter-vocalic transition.

### 1.1. Articulation of geminates

Phonological quantity contrast used in many languages is phonetically realized mostly by differences in acoustic duration of vowels or consonants.

Investigators comparing articulatory patterns associated with the production of phonologically short and long segments report several regularities in kinematic characteristics reflecting this primary phonetic cue. Yet, it should be noted that previous studies use different measurement procedures and linguistic material which can be the cause of variability in the findings. This might suggest that the phonological contrast is not clearly manifested in articulatory patterns across various contexts. Therefore, our study has partly been inspired by these differences.

Studies focused on consonantal gemination show significantly longer durations of articulatory closure for geminates compared to singletons (e.g., Löfqvist, 2006, 2007 for Japanese stops, sonorants, and fricatives; Bouarourou et al., 2008 for Tarifit Berber stops and fricatives; Ridouane, 2007 for Tashlhiyt Berber stops; Zeroual et al., 2008 for Moroccan Arabic stops). The movements of relevant articulators to and from the constriction targets for geminates have longer durations than for singletons (Fivela et al., 2007 for Italian stops and sonorants; Zeroual et al., 2008 for Moroccan Arabic stops; Šimko et al., 2014 for Finnish stops). Also, quantity influences spatial extent of articulatory movements in the expected direction. Larger lip closing and opening movements were found for geminate bilabials compared to their singleton counterparts for example for Italian and Japanese stops and sonorants and for Finnish stops (Fivela & Zmarich, 2005; Löfqvist, 2005; O'Dell et al., 2011). Velocity profiles of relevant articulators exhibit more varied patterns regarding the quantity contrast. Several researchers have reported lower articulatory velocity for long segments than for short ones (e.g., Smith, 1995 and Löfqvist, 2005 for Japanese stops, sonorants, and fricatives; Šimko et al., 2014 for Finnish stop consonants), but others have found significantly higher velocity for geminates than singletons (Löfqvist, 2005 for Swedish and Bouarourou et al., 2008 for Tarifit Berber stops and fricatives). In addition, no significant effect of quantity on movement velocity was reported in several other studies (e.g., Fivela et al., 2007 for Italian stops and sonorants and Zeroual et al., 2008 for Moroccan Arabic stops).

Phonological quantity of a consonantal segment also affects kinematic properties of neighboring gestures. In particular, the duration of articulatory transition between vowels flanking a consonant tends to be longer when the consonant is geminate compared to singleton (Löfqvist, 2006 and Smith, 1995 for Japanese stops and sonorants; Šimko et al., 2014 for Finnish stops [see also an acoustic study by Lehtonen, 1970]; Fivela et al., 2007 for Italian stops and sonorants, but see also Smith, 1995 with different results). Studies offer mixed results regarding the influence of consonantal quantity on spatial extent of the inter-vocalic gesture: No effect for Japanese sonorants (Löfqvist, 2006; but generally greater extent in geminate context for stops and fricatives reported in Löfqvist, 2007) and expansion for geminates in Finnish stops (Šimko et al., 2014). These three studies also suggest slower transition gestures in geminate context.

Vowel quantity also influences kinematic characteristics of gestures participating in the vowel production. Quantity has a significant effect on the durations of the gestures from a constriction preceding and to a constriction following the vowels with shorter durations for the short vowels (Beňuš, 2011 for Slovak; Hertrich & Ackermann, 1997 and Kroos et al., 1997 for German lax/tense contrast). The peak velocity of both preceding and following gestures decreased with quantity; while for German lax/tense contrast the effect of quantity on following gestures is minimal (Hertrich & Ackermann, 1997) for Slovak short/long distinction—generally not accompanied by quality differences—the effect is reported to be greater for the following than the preceding gesture (Beňuš, 2011). The spatial extent of the flanking gestures and of the lingual gestures associated with vowel production is greater for long compared to short vowels (Beňuš, 2011). Interestingly, Hoole and Mooshammer (2002) report a greater spatial extent of the lingual gestures for tense German vowels compared to their lax cognates but only for central and back vowels; quantity does not influence the extent for more articulatorily constrained front vowels.

A consonant, singleton or geminate, is generally coproduced with a following vowel. In the literature, two accounts for coproduction have been suggested. Öhman (1966) claims that vowels are produced continuously, i.e., movements from the first vowel to the following one are produced as diphthongal and consonantal gesture is superimposed on that trajectory. Recent coproduction theory by Fowler and Saltzman (1993) proposes

a view according to which vowels are produced as separate gestures and the gestures for vowels and consonants overlap temporally (see also Fowler, 1980). Using the notion of gestural activation waves, Fowler and Saltzman (1993) show how gestures are coproduced in speech—each gesture influences vocal tract shape smoothly as there are gradual implementation (anticipatory field of coarticulation) and relaxation (carryover field of coarticulation) phases during the activation of a gesture. These phases overlap with adjacent gestures, i.e., they are coproduced. The amount of overlap depends on the extent to which articulators are shared (the degree of spatial overlap). In cases where all articulators are shared, the gestures are competing for the control of articulators during periods of coproduction, which leads to a process of intergestural blending. Thus, context-sensitivity is seen as a consequence of the time courses of the activation waves of gestures and the manner of blending.

Somewhat consistently with the aforementioned theories, Smith (1995) presents two models for the temporal coordination between vowels and consonants. According to the “combined vowel-and-consonant timing” model, vowels and consonants are mutually coordinated, and thus, vowels are affected by the temporal properties of consonants. Smith (1995) showed that this is the case in Japanese (mora-timed language), where the differences in consonant duration influence the timing between the vowel gestures. The “vowel-to-vowel timing” model supports Öhman’s (1966) theory of continuous vowel production, and according to this, vowels are not affected by the properties or the number of consonants. The results of Smith’s (1995) study showed that in Italian (a syllable-timed language) the timing of tongue movements for vowels remains the same regardless of the length of an intervocalic consonant.

More precisely, the gesture associated with the consonant, e.g., lip closing gesture for a bilabial, starts around the same time as the intervocalic transition, i.e., the tongue movement towards the constellation appropriate for the vowel acoustically succeeding the consonant (see e.g., Öhman, 1966). The details of relative timing of the onsets of these two parallel gestures, however, depend on articulatory characteristics of the vowels involved in the intervocalic transition. For example, the transition movement tends to start relatively earlier with respect to the lip closing one when the transition is /a-i/ as opposed to /i-a/ (Löfqvist & Gracco, 1999; Šimko & Cummins, 2009, 2010). The relative timing is also influenced by the preceding consonant: The lip closure onset for a bilabial is delayed relative to the onset of intervocalic transition if the preceding consonant is a homorganic bilabial compared to a non-homorganic one, e.g., /t/ (Šimko et al., 2014; see also a discussion of gestural crowding in Beňuš & Šimko, 2014).

Directly relevant for the present study is a finding reported by Šimko et al., (2014) about an influence of gemination on the intergestural coordination. In the material investigated—Finnish stimuli /paCi/, /piCa/, /taCi/, and /tiCa/ where C is either a singleton or geminate /p/—the duration of the interval between the onsets of lip closing for C and coproduced intervocalic transition was in addition to the aforementioned contextual factors also influenced by the quantity of C. Namely, the lip closing movement started earlier relative to the lingual one when C was a geminate than when it was a singleton. Additionally, the authors report an interaction between context-sensitivity of intergestural timing and gemination in Finnish: The effect of gemination was weaker in the /i-a/ than in the /a-i/-context. Furthermore, in the /i-a/-stimuli, the effect of gemination was still further diminished by the preceding /p/ to the extent that for several speakers the lip closing gesture actually started earlier (relative to the lingual gesture) for singletons than for geminates.

In the current paper, we will study these effects in more detail. In addition to the inter-gestural coordination, we will investigate the interaction between the context and

quantity for other kinematic measures such as duration, spatial extent, and peak velocity of individual gestures. Further, in addition to consonant quantity, we include a variation in vowel quantity. Most importantly, we will study these phenomena on a three-way quantity contrast in one of the very few languages that include such a system, Estonian.

### 1.2. The Estonian three-way quantity

The three-way quantity is the most studied feature of Estonian word prosody, and has been investigated in numerous acoustic and perception studies (e.g., Lippus et al., 2009, 2011; Lippus et al., 2013; Lehiste, 2003; Eek & Meister, 1997, 2003). In comparison, articulatory studies of Estonian quantity are relatively rare.

In a number of studies (e.g., Lippus et al., 2013; Lehiste, 2003; Eek & Meister, 1997) it has been shown that the domain of quantity in Estonian is a left-headed (i.e., stress on the first syllable) disyllabic foot, where stressed syllables can have three degrees of quantity based on syllable length (short and long) and weight (light and heavy): A Q1 syllable is short and light, a Q2 syllable long and light, and a Q3 syllable long and heavy (Viitso, 2003, pp. 11–13). Incorporating all segmental durations within the foot, the quantity opposition between short (Q1), long (Q2), and overlong (Q3) feet is realized by the stressed-to-unstressed syllable rhyme duration ratio and some additional features (namely the pitch movement and vocalic quality). As the unstressed syllables cannot have distinctive length oppositions independent of the stressed syllables, their duration lengthens or shortens compensatorily in the opposite direction of the stressed syllable.

While on the foot level there is a three-way opposition of stressed-to-unstressed syllable length, the stressed syllable rhyme duration is achieved by combining vowel and consonant length. The three-way contrast can be accomplished by lengthening the stressed vowel, the intervocalic consonant, or both, enabling minimal septets of C1V1C2V2-sequences based on segmental duration: Q1—short V1 short C2 [sate] ‘fall-out, sg. nom.’—Q2—long V1 short C2 [sa:te] ‘broadcast, sg. nom.’, short V1 long C2 [satte] ‘fall, sg. nom.’, long V1 long C2 [sa:tte] ‘get, pl. 2<sup>nd</sup> pers.’—Q3—overlong V1 short C2 [sa::te] ‘haystack, pl. part.’, short V1 overlong C2 [sat:te] ‘fall, sg. gen.’, long V1 overlong C2 [sa:t:te] ‘broadcast, sg. gen.’ Regardless of the segmental structure of the stressed syllable, the quantity distinction is best described by the duration ratio of the first and the second syllable rhymes within the foot. The ratio is robustly 2/3 in Q1, 3/2 in Q2, and 2/1 in Q3 (Lehiste, 2003; Lippus et al., 2013).

Additionally, quantity is marked by pitch movement: In Q1 and Q2 words, the peak of the pitch contour is at the end of the first syllable, while in Q3 it is located at the beginning of the first syllable (Lehiste, 2003; Lippus et al., 2011; Lippus et al., 2013). Vocalic quality has also been shown to vary along with the quantity (Eek & Meister, 1998), and it is significant for distinguishing Q1 from Q2 and Q3 (Lippus et al., 2013).

From the phonological point of view, the Estonian quantity distinction has also been a subject of discussion regarding moraic interpretation. Fitting the Estonian three-way quantity contrast within the frames of moraic theory appears to be a challenging task. Provided that a maximum number of moras a syllable can consist of is two, the ternary contrast should be fitted into a bimoraic space, i.e., trimoraic syllables are not allowed. Thus, the main issue regarding Estonian quantity lies in the representation of long and overlong syllables. Hint (1973) and Prince (1980) have treated Q1 and Q2 syllables as monomoraic and Q3 as bimoraic, but Prillop (2013) points out the contradiction of short and long vowels which would both be monomoraic in that case.

The easiest way of describing Estonian quantities would be to use trimoraic syllables, as has been done by Hayes (1995), who links stressed Q1, Q2, and Q3 syllables with one,

two, and three moras respectively. Alternatively, Bye (1997) has also used mono-, bi-, and trimoraic syllables, but for keeping the maximum number of two moras in a syllable, he suggests using extrasyllabic moras according to which the third mora of Q3 is either a free mora (does not belong to the syllable) or part of a degenerate syllable. Similarly, Eek and Meister (2004, pp. 351–352) propose that Q3 could constitute a syllable followed by a degenerate syllable which together form a separate foot. However, these theoretical approaches use trimoraic vowels and are not fully supported by phonetic measurements as the difference between trimoraic vowels in Q3 and monomoraic vowels in Q1 is too big, as is the difference between Q2 and Q3 syllables (Prillop, 2015).

Another possibility suggested by Eek and Meister (2004) is that the Q1 syllable is monomoraic and the Q2 and Q3 syllables are bimoraic with the difference between the latter two being in the mora sharing of Q3 syllable nucleus and the coda, whereas this does not happen in Q2 syllable. While Eek and Meister considered the second component of a diphthong as the coda, this approach does not fit well in the case of long vowels that would still be trimoraic (Prillop, 2013).

According to Prillop (2015) the representation of three-way distinction of consonants would be better explained by the possibility that coda consonants can either be linked to a mora or not. In Q2 syllable, coda consonant could not be considered moraic, while the coda consonant could be linked to a mora in Q3. This representation is not suitable for diphthongs (the second part of a diphthong would have to be reduced) nor for long vowels (it would assume monomoraic long vowels).

Considering previously discussed issues concerning the representation of Estonian three-way quantity distinction from the perspective of moraic theory, the question arises whether using a unit of mora is adequate in Estonian. Lehiste (1990) has compared Estonian and Japanese in terms of the temporal structures of poetry and concluded that in Estonian mora is not an isochronous unit of timing as it is in Japanese. She points out that timing in Estonian is connected to syllables, and that the basic rhythmic unit is a disyllabic sequence. This hypothesis is also supported by numerous phonetic findings about Estonian quantity.

The latest phonological descriptions of Estonian three-way quantity system (Prillop, 2013, 2015) have also taken phonetic facts into account and have been based on the approach of the ternary system being manifested in a disyllabic foot as a whole. Feet are disyllabic for Q1 and Q2, but the stressed syllable of Q3 constitutes a foot alone, so the following unstressed syllable is reduced because it does not belong to the foot. The final-lengthening at the end of the Q3 syllable and at the end of the Q1 foot is caused by the strong mora. Incorporating the Iambic-Trochaic Law (Hayes, 1995) Prillop (2013) expands on Kager's (1993) idea of weak and strong moras and suggests the opposition of weak and strong moras for Estonian. Strong moras are phonetically described with longer or more intense sounds. Therefore, Q1 and Q2 feet have a strong mora in the second syllable (where the vowel is lengthened) and Q3 feet have a strong mora in the first syllable. Stressed Q1 syllable is monomoraic and stressed Q2 and Q3 syllables are both bimoraic, whereas the difference is in the location of the strong mora.

### **1.3. Previous articulatory studies of Estonian**

Research done on articulatory aspects of Estonian quantity is relatively scarce and most of it dates from the 1960s and 1970s. In his pioneering articulatory study of quantity on Estonian vowels, Liiv (1961) found a substantial difference in the spatial extent of gestures between short and long vowels on the one hand and overlong ones on the other (although some methodological shortcomings have later been identified; Eek, 2008, pp. 67).

Eek (1970a, 1970b, 1970c, 1971a, 1971b, 1971c) studied the articulation of Estonian sonorant consonants (/n, n<sup>j</sup>, l, l<sup>j</sup>, m, r/) in three quantity degrees using different methods (X-ray, palatography, cinefluorography, and filming of lip articulation). The sonorants occurred in sentence-initial and in isolated words in the /a-a/-context, and also in a word-final position. His studies show that contact area and stiffness (estimated from cinefluorographic film) of articulatory movement increases with longer duration of the consonant, while the lip aperture is narrowing. However, there is no clear three-way distinction in the articulatory characteristics. In the case of /n/ and /l/ in Q2 and Q3, the contact area between the tongue and the palate is larger than in Q1. Linguopalatal contact is also wider for /r/ in higher quantity degrees. There is more variation in lip aperture; for /n/, with an increasing duration of the consonant it is either decreasing (3 subjects out of 5) or increasing (2 out of 5); for /l/ the lip aperture estimates in Q1 and Q2 are of similar length, but for Q3 there is more variation between subjects. With increasing quantity of /r/ Eek reports a regular narrowing of the lips. When making comparisons between consonants, differences between sonorant groups can be found. In the same quantity degree there is always a longer alveolar contact for /l/ than for /n/. The biggest area of lateral contact is found for /r/. The lip aperture is wider for /l/ and /r/ than for /n/.

The speed of the articulatory movements from a vowel to a following consonant is the slowest in Q1 and the fastest in Q3 (for both /n/ and /l/). Faster speed for Q3 sonorant shows greater muscular effort. This is also the case for the bilabial nasal /m/, which is pronounced more tensely in Q3 than in Q1 and Q2.

In general, Eek concludes that while the difference between Q1 and Q2 is primarily in the durational characteristics, Q3 is realized by significantly faster articulatory movements. In overlong consonants the first, syllable final component is produced with greater muscular effort, so the beginning of the consonant is tense. In the case of a long consonant, the syllable final component is lax. This is in accordance with the idea of virtual target (Löfqvist & Gracco, 1997; Löfqvist, 2005) suggesting that the position of the virtual target for the lips is varied in order to control the duration of closure/constriction, meaning higher target and bigger amplitude for long consonants. Since displacement is claimed to strongly correlate with velocity, long consonants should also have higher peak velocity. While Eek's findings seem to support the hypothesis for both the displacement and velocity, Löfqvist (2005) found no differences in velocity when examining the production of Japanese and Swedish labial stop and fricative consonants varying the duration of the oral closure/constriction (long and short). Thus, Löfqvist concluded that for keeping the contact between lips longer, speakers vary both the position of the virtual target and its timing by modifying the deceleration of the lower lip movement.

Lehiste et al., (1973) used electromyography (EMG) for examining the activity of the orbicularis oris muscle during the production of bilabial single and geminate Estonian consonants in intervocalic position and at word boundaries. The results showed a difference in progressive closure duration of bilabials in all three quantity degrees. The authors also bring out the two-phase theory of geminate production according to which the first phase includes a syllable final occurrence of the consonant; the second phase is rearticulated and it starts the next syllable (forming a syllable-initial consonant). The analysis of EMG data suggests that there is a difference between long and overlong consonants, which exhibited a 'rearticulation' in their production, i.e., EMG signal revealed two successive peaks for a long consonant.

In a recent palatographic study on Estonian sonorant consonants, Meister and Werner (2015) compare the results obtained using a contemporary EPG system with Eek's findings from the 1970s. They conclude that the articulatory characteristics of sonorants have stayed the same during the years. Similar to Eek, the authors confirm that higher

quantity degrees are accompanied with larger contact area of the articulators involved. Nevertheless, there are some differences between the studies. The new results show a different size of the contact areas for /n/ in Q2 opposed to Q1 and Q3, as compared to Eek who claims that Q1 differs from the other quantities. Meister and Werner's (2015) data also show a three-way distinction for the palatalized /l/.

#### **1.4. Context-dependency and research questions**

The syllable rhyme duration ratio—a traditional and robust way of describing the quantity distinction in Estonian—has been typically ascertained within similar segmental context (e.g., Eek & Meister, 2003; Lippus et al., 2009) or by ignoring the segmental context by randomly pooling different phoneme combinations together (e.g., Lippus et al., 2013). At the same time, the intrinsic properties of sound segments, claimed to have rather a psychophysiological than language-specific nature (Lehiste, 1970), have been shown to have non-linear effect on the quantity-related duration variability in Estonian. Eek (1974) has demonstrated that while /t/ is shorter than /p/, the ratio between the segments in Q1, Q2, and Q3 is also different: For /p/ it is 1 : 1.3 : 2.1, and for /t/ 1 : 2.0 : 2.9. As for perception, consonant context does not have a significant effect on the following vowel, but vowel quality can shift the perceived short-long category boundary earlier for /i/ than for /a/ (Meister et al., 2011).

Lippus and Šimko (2015) investigated context-sensitivity using the acoustic data of the material described in Section 2 of the current paper (see below). They show a strong influence of the intrinsic characteristics of segments on the temporal patterns of quantity combinations. In all quantity degrees /p/ is longer than /t/. Regarding vowel context, /a/ is longer than /i/, but there is a non-linear pattern associated with vowel quantity of the first syllable. The interaction between vocalic and consonant context somewhat neutralizes the word-initial consonant context, but vocalic context effect gets stronger in the interaction with the consonant. Importantly, the durational differences marking the quantity contrast are in some cases not robust enough to make these contextual effects negligible; the contextual effects need to be taken into account by the listeners and 'compensated away' in order to correctly assess the quantity pattern of a given stimulus. These results raise a question about the extent that individual influences on the rendition and parsing of various speech items varying in segmental context and quantity can be interpreted in terms of a system of discrete 'phonological' contrasts.

In this work we extend these investigations of context-dependency of quantity marking in Estonian to its articulatory aspects. We ask how relationships among kinematic characteristics reflect different quantity patterns in Estonian in different segmental contexts:

- (1) Is the Estonian three-way quantity contrast manifested by a corresponding three-way distinction in articulatory characteristics?
- (2) What are the articulatory correlates of quantity in terms of adjustments of gestural kinematic characteristics and an intergestural coordination measure to different quantity patterns?
- (3) How does segmental context influence articulatory characteristics and coordination and, mainly, to what extent does its influence interact with quantity correlates?

Guided by these questions, our primary aim is to evaluate the robustness of articulatory characteristics of the phonological quantity contrast with respect to the variation of speech material. To address these issues, we present an analysis of articulatory material

consisting of renditions of CVpV stimuli in all 7 contrastive patterns of the Estonian quantity system. The initial C is a bilabial /p/ or an alveolar /t/ and the vowels are /a/ or /i/ (only the two conditions with different vowels within a stimulus were recorded). These variables—stimulus initial consonant and vowels flanking the stimulus internal consonant /p/—present the segmental context relevant for question (2).

Three articulatory gestures are included in the analysis: The lip opening and lip closing gestures for stimulus internal /p/ and the intervocalic transition gesture between the flanking vowels. These gestures were chosen as they are deemed to appropriately represent the articulatory events related to quantity contrast and, unlike the other remaining lingual gestures, do not *a priori* involve immediate interaction with multiple contextual elements. That is, while, for example, intervocalic transition is primarily coproduced with /p/, it can be fully characterized by a movement from /a/ to /i/ or from /i/ to /a/. On the other hand, the tongue movement towards the first vowel is coproduced with either /p/ or /t/ as well represents one of the two possible transitions between vocalic segments.

For each of these gestures, we only report data pertaining to three ‘raw’ kinematic characteristics: Duration and spatial extent of each gesture, and peak velocity. We refrain here from including ‘derived’ measures such as stiffness estimates, c-parameter (Ostry & Munhall, 1985) or peak-to-peak ratio (Harrington et al., 1995), etc. Apart from parsimony’s sake, the main reason for this decision is that we do not at this stage commit to necessary theoretical underpinning of gestural dynamics associated with these measures, as this work is one of the first large-scale investigation of articulation in the Estonian three-way quantity system. We plan to address the questions of intergestural stability and other dynamic characteristics of articulation in subsequent work.

## 2. Materials and methods

### 2.1. Speech material

The test subjects were 4 native Estonian speakers (2 female and 2 male, mean age 41, ranging from 34 to 57). The speakers had no reported speech or language processing disorders. The recording of the speech material was guided by the ethical guidelines of the University of Helsinki.

The audio signal, and the tongue (TB), jaw, and lip (UL and LL) movements were recorded using electromagnetic articulography (EMA, AG500, Carstens Medizinelektronik) at the University of Helsinki. Subjects produced a CVpV sequence in all possible quantity combinations. In the stimuli, the initial consonant and vocalic context were varied, the initial consonant was either /p/ or /t/, and the vowel context either /a-i/ or /i-a/. The stimuli thus differed in their quantity patterns (7 possibilities), initial consonant context (2) and vocalic context (2), resulting in 28 test stimuli listed in **Table 1**. As seen in the table, the 7 quantity patterns of Estonian arise through gradual three-way lengthening of a vowel (V1) and/or a consonant (C2). The Estonian three-way quantity contrast is thus realized by increasing *consonant length* (Q1 vs. Q2-CL vs. Q3-CL), *vowel length* (Q1 vs. Q2-VL vs. Q3-VL), or both *vowel and consonant length* (Q1 vs. Q2-VCL vs. Q3-VCL).

The stimuli were presented to the participants in standard Estonian orthography. As Estonian orthography does not capture all quantity-related contrasts (e.g., distinction between long and overlong vowels) and only seven of the test words are meaningful words in Estonian (marked with \* in **Table 1**: *tiba* [tipa] ‘a bit’; *papi* [pap:i] ‘old man’/‘priest, sg. gen.’/‘cardboard, gen.’; *tapi* [tap:i] ‘dovetail joint, sg. gen.’; *tipa* [tip:a] part of *tipa-tapa*, ‘walking lightly’; *pappi* [pap:i] ‘priest, sg. part.’/‘cardboard, part.’; *tappi* [tap:i] ‘dovetail joint, sg. gen.’; *tiiba* [ti:ipa] ‘wing, sg. part.’), each stimulus was presented together with a sentence where a segmentally similar word with the same quantity degree was used, e.g., [pa:pi] (Q3-VL) *Töömehel on vaja saagi*. ‘The workman needs **the saw**.’; [tipa] (Q1)

**Table 1:** Stimuli for each of the seven possible quantity combinations.

	Q1			Q2			Q3			
		CL	VL	VCL	CL	VL	VCL	CL	VL	VCL
papi		pap:i*	pa:pi	pa:p:i	pap:i*	pa::pi	pa:p:i			
pipa		pip:a	pi:pa	pi:p:a	pip::a	pi::pa	pi:p:a			
tapi		tap:i*	ta:pi	ta:p:i	tap:i*	ta::pi	ta:p:i			
tipa*		tip:a*	ti:pa	ti:p:a	tip::a	ti::pa*	ti:p:a			

\* Marks the combinations that are meaningful words in Estonian.

*Tulin tība hilja.* ‘I arrived a bit too late.’ The subjects were instructed to repeat each stimulus about ten times without pauses between the repetitions;<sup>1</sup> the order of the stimuli was randomized. To eliminate the boundary lengthening effects, the first and the last repetitions were excluded from the analysis. Between 118 and 129 tokens were collected for each stimulus.

## 2.2. Post-processing

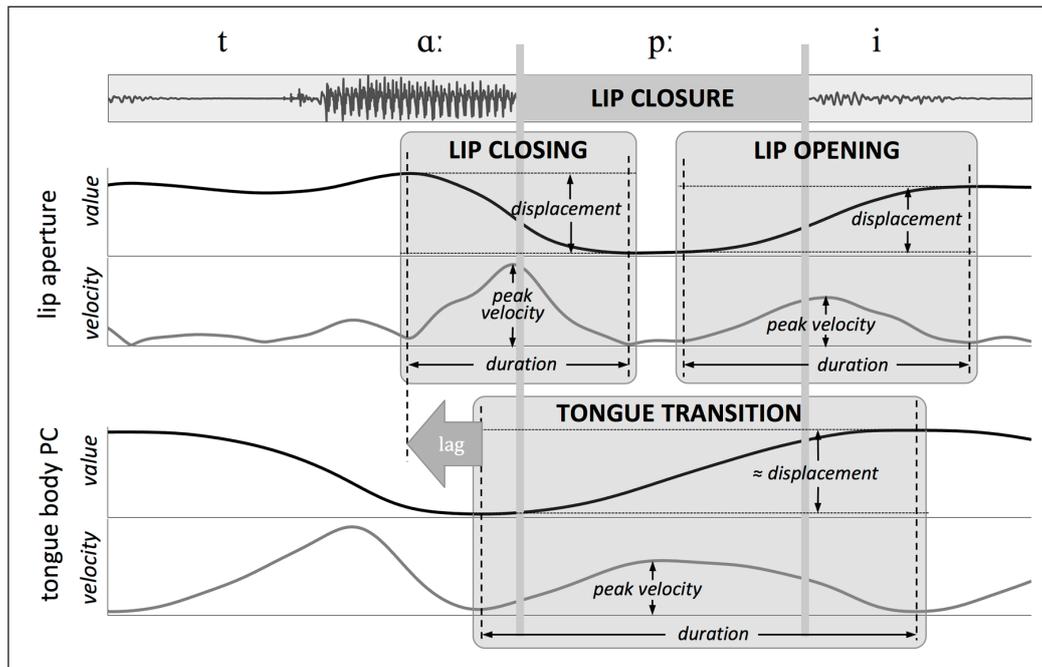
The relevant receivers were placed on the midsagittal plane of the vocal tract; the lip sensors (UL and LL) were placed above and below the vermilion border of the upper and lower lip, respectively, the jaw sensor below the lower incisors, and the tongue sensor (TB) in a middle portion (blade) of the tongue.

The articulatory trajectory signals, recorded at a sampling rate of 200 Hz, were processed using the Tapad system (Hoole & Zierdt, 2010) and subsequently corrected for head movements, smoothed by an 8-point Bartlett window and up-sampled to 1000 Hz using cubic spline interpolation. Only two dimensions—projections of the sensors’ trajectories to mid-sagittal—were used for articulatory labeling.

Onsets and offsets of the bilabial closure of stimulus internal /p/ were manually labeled by an experienced annotator, using the acoustic signal recorded synchronously with the articulatory data. The acoustic closure onsets and offsets were used as anchors for semi-automatic articulatory labeling implemented in Matlab used to identify the onsets and offsets and peak tangential velocity of the three gestures of interest (see **Figure 1**). A relative lip position measure, i.e., the Euclidean distance between the lip receivers, was calculated from the position of the upper lip with respect to the lower lip sensor. The onset and offset of the lip closure movement for the bilabial were found as the local minimum of tangential velocity of this ‘lip aperture’ measure preceding and following the acoustic lip closure onset, respectively (see **Figure 1**). Similarly, the onsets and offsets of the lip opening gesture of the closing movement were identified as local minima of tangential velocity ‘surrounding’ the acoustic offset of the lip closure. The onsets and offsets of the inter-vocalic transition gesture between target positions corresponding to the vowels flanking the /p/ were again identified as the local minima of tangential velocity of TB sensor. All automatically determined labels were checked by an annotator and corrected when necessary.

For each gesture, *duration* is a length of the temporal interval between the articulatory onset and offset of the gesture. *Peak velocity* is a maximal value of the tangential velocity signal during the interval. Finally, *displacement* was computed as a Euclidean distance

<sup>1</sup> Our choice of using repetition was motivated by two reasons. Given the number of stimuli, the more usual approach to embed each stimulus in a carrier phrase would result in either too few stimuli or unduly long recording sessions. Also, in a similar study of Šimko et al., (2014) on Finnish geminates where both repetition and carrier sentence approach was used in parallel, the authors found no significant differences in the measures of interest (the same as in the current study) between the conditions.



**Figure 1:** Illustration of articulatory signal labeling. From top to bottom: Acoustic waveform with indicated manually annotated lip closure, ‘lip aperture’ value and tangential velocity and tongue body sensor value and tangential velocity (as both ‘lip aperture’ and TB sensor signals are 2-dimensional, distance between the lip sensors and principal component of TB trajectory are plotted, respectively). For all three gestures of interest (shaded rectangles), the values of articulatory measures used in this work are depicted (duration, displacement, and peak velocity). Also, ‘lag’ label depicts the lip-tongue coordination measure.

between the relevant position measure (position of the TB sensor, value of the derived ‘lip aperture’ measure) at the onset and offset of the gesture.

In addition, the *lip-tongue coordination* measure was computed as the onset time of the vowel transition movement minus the onset time of the bilabial closing movement (see Löfqvist & Gracco, 1999; Šimko & Cummins, 2010, 2011; Šimko et al., 2014).

### 2.3. Statistical analysis

The aim of the statistical evaluation of the recorded material was to assess whether and to what extent the variations in terms of quantity and segmental context influence the articulatory measures under investigation. Importantly, we wanted to verify whether these influences can be meaningfully interpreted in terms of discrete, phonological contrasts (e.g., word-initial /p/ vs. /t/, everything else being equal), or whether the observed articulatory patterns arise through complex interactions that, in essence, render each individual test word a separate articulatory task.

Addressing these questions, we used mixed effect modeling with every articulatory measure as a dependent variable: Duration, displacement, and peak velocity of each of the three articulatory gestures—lip closing, lip opening, tongue transition—and the lip-tongue coordination measure, thus fitting 10 separate models. Word-initial consonantal context (levels: /p/ and /t/), vocalic context (levels: /a-i/ and /i-a/), and quantity (seven quantity combinations, cf. **Table 1**) were used as fixed factors. Also, to verify the effect of the repetition of items in our recording paradigm, we include Repetition number as a fixed effect. Therefore, in final models the influence of the four explanatory variables and their interactions were tested with adjusting the differences in the three main fixed predictors for each speaker by adding random intercepts and slopes for C1-context, V1-context, and

quantity by test subject. Adding random slopes allows subjects to vary with respect to the aforementioned three effects, that is, it avoids the assumption that the effects would be identical across speakers (cf. Barr et al., 2013; Baayen & Milin, 2010). Item was not added in the models as random factor, since the individual characteristics of consonant context, vocalic context, and quantity fully describe the items observed, i.e., the items systematically vary in these characteristics and no other.

Each model was fitted using lmer function of lme4 package in R. The first and the last repetition of each item in a block was excluded from the data set. The models were not tested in an incremental fashion to evaluate the significance of each of the fixed factors for every dependent variable, since we are not primarily interested in a mere presence or absence of significant effect of these variables but we use multiple comparison technique to quantitatively evaluate the differences made individual contrasts (and their direction and significance). For the sake of clarity and systematic analysis we use the same models for each dependent variable and report the results of pairwise comparisons.

Statistical models were analyzed using a Tukey HSD multiple comparison technique evaluating the statistical significance of differences between various estimates being different from zero. The pairs of items differing in a single ‘phonological’ parameter, i.e., Consonant Context, Vocalic Context, and Quantity were included in the set of linear hypotheses to be tested, resulting in 64 hypotheses with *p*-values corrected for multiple testing. The estimates for the second repetition in each block were used for comparisons. The glht function of multcomp package of R was used to quantify the significance of the hypotheses.

The resulting mixed effect models as well as the results of the post-hoc pairwise comparisons are presented in Appendices A and B, respectively.

### 3. Results

#### 3.1. Acoustic duration of intervocalic stop

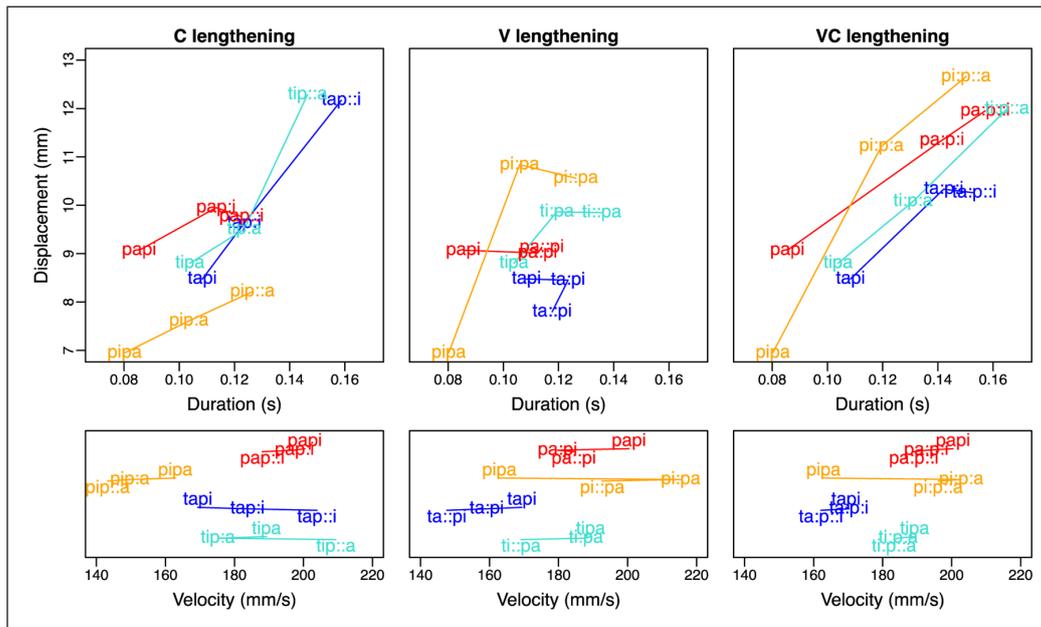
The acoustic duration of closure for the intervocalic bilabial stop /p/ is influenced by the consonant quantity (CL Effect) with consistent increase in closure duration with higher quantity degree showing a three-way distinction ( $p < 0.001$  for all of the comparisons; see Appendices A11 and B11 for full results of statistical modeling and the x-axis of **Figure 5** for the duration of oral closure). While vowel lengthening (VL Effect) has no significant influence on the acoustic duration of the intervocalic stop, the stimuli with both vowel and consonant lengthening (VCL Effect) reflect the ternary effect of consonant quantity ( $p < 0.001$  for all comparisons, except  $p < 0.01$  for Q1 and Q2 stimuli *pipa—pi:p:a*, *tapi—ta:p:i* *tipa—ti:p:a*). Regarding vocalic (/a-i/ vs. /i-a/) and consonantal (/p/ vs. /t/) context effects, basically none of the differences are significant; the duration of the intervocalic stop is longer for *ta::pi* compared to *pa::pi* ( $p < 0.05$ ).

#### 3.2. Lip closing

The results for the lip closing gesture are shown in **Figure 2** with the duration (x-axis) and displacement (y-axis) depicted in the upper three panels and the measures for the peak velocity at the bottom.

##### 3.2.1. Duration

As seen in the upper leftmost panel in **Figure 2**, duration of the lip closing gesture generally increases with increasing consonant length (CL Effect). According to the model prediction, the tree-way effect emerges for the stimuli with /i-a/ vocalic context ( $p < 0.001$ ;  $p < 0.01$  for *pip:a—pip::a*;  $p < 0.05$  for *tipa—tip:a* and *tip:a—tip::a*). With /a-i/-context, Q1 differs from Q2 and Q3 in the case of /p/-initial stimuli, and Q1 and Q2 from Q3 in the case of /t/-initial stimuli ( $p < 0.001$  for all comparisons).



**Figure 2:** Lip closing gesture. The kinematic characteristics are estimated for the duration, displacement (upper panels), and the peak velocity (bottom) of the articulatory movements. The panels show the stimuli with the CL Effect (left), VL Effect (mid), and VCL Effect (right).

Bigger vowel quantity is also accompanied with longer lip closing gestures. A two-way effect of vocalic length (VL Effect, see the x-axis of the upper middle panel in **Figure 2**) on the duration of the lip closing gesture is significant for /p/-stimuli with an intervocalic singleton consonant differing only in VL ( $p < 0.01$  for Q1 vs. Q2 papi—pa:pi and pipa—pi:pa;  $p < 0.001$  for Q1 vs. Q3 papi—pa::pi and pipa—pi::pa). Q2 and Q3 counterparts are not significantly different. Regarding the /t/-stimuli, VL Effect only elicits differences for the tipa—ti::pa pair ( $p < 0.001$ ) with a longer closing gesture for ti::pa.

When the quantity contrast is realized by both vowel and consonant length (CVCV—CV:C:V—CV:C::V) as shown in the right-hand panel in **Figure 2**, the patterns of the lip closing gesture duration follow the ones for CL Effect: The three-way effect is predicted for /i-a/-context ( $p < 0.001$ ;  $p < 0.01$  for tipa—ti:p:a) and a two-way VCL Effect of Q1 vs. Q2 and Q3 for /a-i/-context ( $p < 0.001$ ).

The vocalic context (/a-i/ vs /i-a/) does not influence the lip closing gesture duration in a consistent way and, in fact, the differences are significant for two pairs of stimuli. The closing gesture is longer for pa:p:i than for pi:p:a ( $p < 0.001$ ), and longer for ti::pa compared to ta::pi ( $p < 0.01$ ). A more consistent pattern emerges regarding the influence of stimulus initial consonant. For each pair of stimuli differing solely in this aspect, the model estimates of duration are greater in /t/- than in the /p/-stimuli, but again, this difference is generally not significant, except for the papi—tapi, pap:i—tap:i, and pipa—tipa pairs ( $p < 0.01$ ).

The repetition of the stimulus showed no significant overall effect on the duration of the lip closing gesture. Compared to the base, Q1 stimulus papi, the effect becomes significant in the interaction with the /t/-context and quantity: The duration is increasing with repetition in the case of Q2 and Q3 stimuli tap:i, ta:p:i, tap::i, ta:p::i ( $p < 0.05$ ).

### 3.2.2. Displacement

The spatial extent of the lip closing gesture is less affected by quantity. CL Effect on the displacement for the stimuli with a short vowel (y-axis in the leftmost plot in **Figure 2**) is significant for the /t/-stimuli forming a two-way effect of Q1 and Q2 vs. Q3 ( $p < 0.001$ )

with greater displacement for the latter. A two-way contrast also marks the VL Effect in the /i-a/-context in the /p/-stimuli (pipa—pi:pa—pi::pa), but the displacement is significantly greater for the stimuli with long and overlong vowels compared to short ones (pipa—pi:pa at  $p < 0.001$ ; pipa—pi::pa at  $p < 0.01$ ).

The same two-way result emerges regarding VCL Effect (shown in the rightmost panel in **Figure 2**) on /p/-initial stimuli, as the displacement of the closing gesture is greater for Q2 and Q3 than Q1 words ( $p < 0.001$ ). For the /t/-initial stimuli the displacement of the gesture is larger for Q2 ta:pi compared to Q1 tapi ( $p < 0.05$ ) and also greater for Q3 ti:p::a compared to Q1 tipa ( $p < 0.001$ ).

There is no clear pattern regarding the influence of consonantal and vocalic contexts on the lip closing gesture displacement. In fact, Vocalic Context Effect elicits almost no significant differences between any of the pairs of stimuli (only for pap:i—pip:a at  $p < 0.05$  with significantly greater displacement for the former, /a-i/-context). Consonant Context Effect shows a significantly greater displacement of the gesture with the word-initial /t/ compared to /p/ for Q3 pairs tap::i—pap::i ( $p < 0.05$ ) and tip::a—pip::a ( $p < 0.001$ ), but not for any other pair.

Repetition has an overall decreasing effect ( $p < 0.01$ ) on the displacement of the closing gesture, i.e., the more times the stimulus has been uttered, the smaller the spatial extent gets. However, using papi (Q1) as a base stimulus, the interactions between stimulus and repetition show some variation. Comparisons exhibit a significant, but small shortening effect on the spatial extent for ta:pi and ti:pa ( $p < 0.05$ ) and an increasing effect for tapi ( $p < 0.05$ ), pa:pi ( $p < 0.01$ ), pa:p::i ( $p < 0.05$ ) stimuli.

### 3.2.3. Peak velocity

The peak velocity of the lip closing gesture is not affected by consonant nor vowel length, and the comparisons also show no differences regarding Vocalic and Consonant Context Effects. This can be seen in the bottom three panels in **Figure 2** where the triplets of the stimuli are more or less clustering together. The only significant differences emerge for stimulus pairs tapi—tap::i, tip:a—tip::a ( $p < 0.05$  for both) with faster gestures for the latter (Q3) counterparts.

The lip closing is significantly faster with /a-i/ than /i-a/ vocalic context for the pap:i—pip:a and pap::i—pip::a stimulus pairs ( $p < 0.05$ ) and faster with the /t/-context compared to the /p/-context for tip::a—pip::a ( $p < 0.01$ ).

The main effect of the Repetition on the peak velocity of the lip closing gesture is significant (at  $p < 0.01$ ) showing that the movements get slower with each repetition. The interactions indicate significantly faster gestures with each repetition for the tapi ( $p < 0.01$ ), pa:p::i, ti:pa ( $p < 0.05$ ) stimuli and slower gestures for ta:pi ( $p < 0.05$ ).

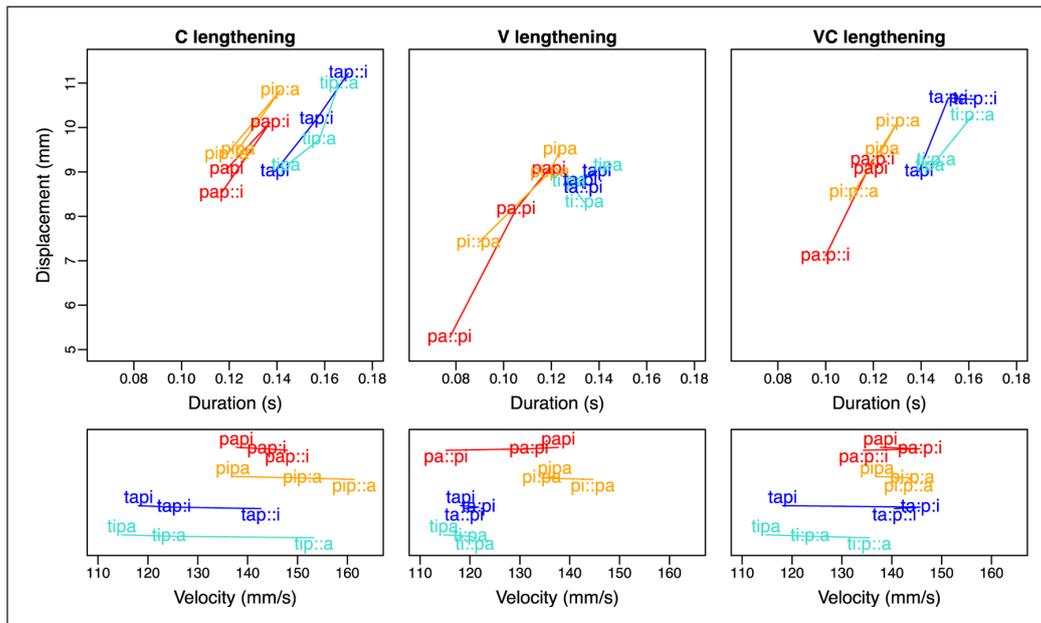
## 3.3. Lip opening

The results for the lip opening gesture are shown in **Figure 3**.

### 3.3.1. Duration

The upper panels in **Figure 3** illustrate the results for the lip opening gesture duration (x-axis). Vowel lengthening (VL Effect, middle panel in **Figure 3**) has a significant effect on the gesture duration in /p/-initial stimuli differentiating Q1 and Q2 from Q3 (at  $p < 0.001$  for all comparisons) with shorter opening with Q3 vowels. CL and VCL Effects show no significant differences between any of the pairs of stimuli ( $p > 0.05$ ), except for Q2 vs. Q3 pa:p:i—pa:p::i ( $p < 0.01$ ) also with shorter opening for Q3 word.

There is no considerable effect of Vocalic Context Effect on the lip opening gesture duration; the opening gesture is significantly longer with the /i-a/ than /a-i/-context in the case of



**Figure 3:** Lip opening gesture. The kinematic characteristics are estimated for the duration, displacement (upper panels), and the peak velocity (bottom) of the articulatory movements. The panels show the stimuli with the CL Effect (left), VL Effect (mid), and VCL Effect (right).

pa:pi—pi:pa ( $p < 0.05$ ) and pa::pi—pi::pa ( $p < 0.01$ ). Regarding Consonant Context Effect, the lip opening gestures are consistently longer in the /t/-context than in the /p/-context with statistically significant differences for all pairs of Q3 stimuli ( $p < 0.001$ ) and Q2 stimuli with /a-i/ vocalic context ( $p < 0.05$  for ta:pi—pa:pi;  $p < 0.01$  for ta:p:i—pa:p:i).

The effect of Repetition on the duration of the lip opening gesture in general is not significant, but gains significance in some interactions with Q3 vowel: The duration of the gesture decreases in pi::pa and increases in ti::pa ( $p < 0.05$ ).

### 3.3.2. Displacement

The results for the displacement of the lip opening gesture (y-axis in **Figure 3**) are reminiscent of the findings for the duration of the opening gesture. In the /p/-context the displacement is smaller in words with the Q3 vowel compared to Q1 and Q2 ( $p < 0.001$  for pa:pi—pa::pi and papi—pa::pi;  $p < 0.01$  for pi:pa—pi::pa and pipa—pi::pa). The VCL Effect is significant only for pa:p:i—pa:p::i ( $p < 0.05$ ) where the opening is smaller for the latter. No other comparisons show significant differences.

As for the effect of Vocalic Context Effect on the displacement of the lip opening, the differences between all pairs of stimuli are non-significant. Consonant Context Effect reveals similar patterns to gesture duration as the displacement tends to be greater in the /t/-context than in the /p/-context, with significant differences for Q3 stimuli ( $p < 0.001$  for tap:i—pap::i, ta:pi—pa::pi, ta:p:i—pa:p::i;  $p < 0.05$  for tip:a—pip::a and ti:p:a—pi:p::a; except  $p > 0.05$  for ti::pa—pi::pa).

The Repetition has a significant ( $p < 0.05$ ) overall decreasing effect on the displacement of the lip opening gesture. The interactions show significant differences comparing the base stimulus Q1 papi with Q3 stimuli. The gesture displacement is increasing with repetition in the case of pa:p:i ( $p < 0.05$ ) and ti::pa ( $p < 0.001$ ), and decreasing in the case of ta:pi and pi::pa ( $p < 0.01$  for both).

### 3.3.3. Peak velocity

The bottom three panels in **Figure 3** depict the results for the peak velocity of the lip opening gesture. The velocity of the gesture is not consistently influenced by quantity. CL

effect has a two-way influence of Q1 and Q2 vs. Q3 on the velocity of the opening gesture in the /t/ + /i-a/-context with faster gestures in Q3 ( $p < 0.001$  for tip:a—tip::a;  $p < 0.01$  for tipa—tip::a). VL Effect differentiates Q1 from Q3 in the stimulus pair papi—pa::pi ( $p < 0.01$ ), where the gesture is faster in the former, Q1 stimulus. Regarding VCL Effect, the opening gesture is faster in Q2 ta:p:i than in Q1 tapi ( $p < 0.01$ ). There is no effect of vocalic nor consonantal context.

Repetition has a decreasing main effect ( $p < 0.05$ ) on the velocity, but the interactions show a significant increase in the case of pa:p:i ( $p < 0.05$ ).

### 3.4. Tongue transition

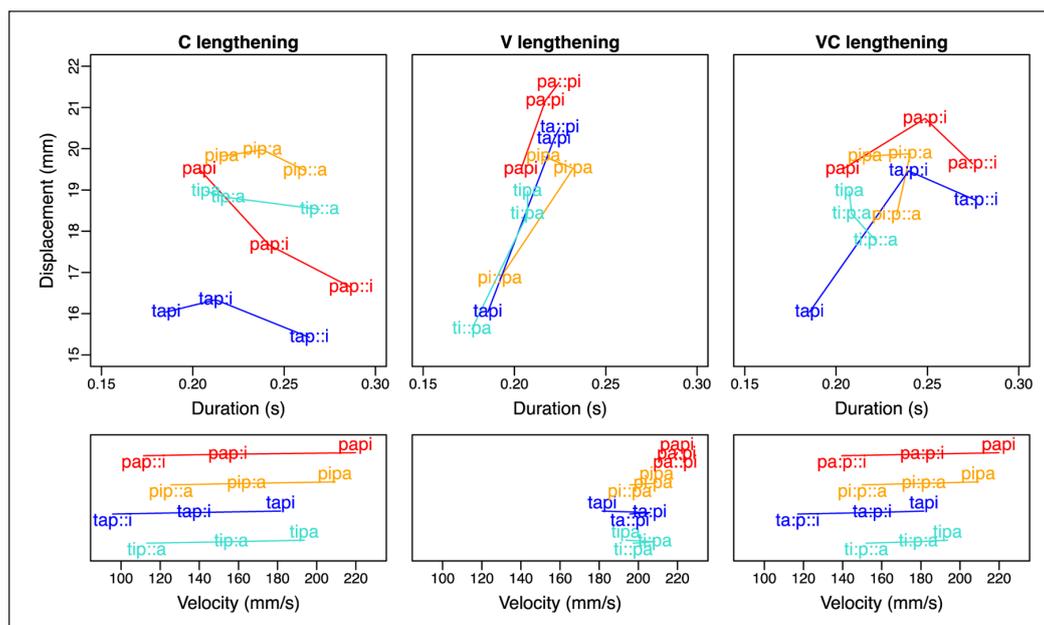
**Figure 4** shows the results for the kinematic characteristics of the lingual gesture of transition between the vowels surrounding the bilabial consonant.

#### 3.4.1. Duration

The duration of the lingual gesture of transition between the vowels flanking the bilabial consonant is influenced by CL Effect with longer gestures for words with greater consonant quantity. As it can be seen from the upper left panel in **Figure 4**, there is a significant difference between Q1 and Q2 vs. Q3 stimuli in words with /a-i/ vocalic context (at  $p < 0.001$  for Q2—Q3 pairs and  $p < 0.01$  for Q1—Q3 pairs). The transition is also significantly longer for tip::a compared to tip:a ( $p < 0.001$ ).

Vowel lengthening effect on gesture duration (x-axis of the upper middle panel in **Figure 4**) also tends to depend on the vocalic context. The effect is clear in /t/-initial words, where with /a-i/-context the transition gesture is longer with Q2 and Q3 compared to Q1 stimuli ( $p < 0.05$  for tapi—ta:pi;  $p < 0.001$  for tapi—ta::pi), but with /i-a/-context it is shorter with Q3 vowels as opposed to Q1 and Q2 ( $p < 0.01$  for ti:pa—ti::pa and  $p < 0.05$  for tipa—ti::pa). The lingual transition between vowels is also shorter with Q3 compared to Q2 stimuli pi:pa—pi::pa ( $p < 0.001$ ).

When combining the influence of both consonant and vowel lengthening, the duration of the lingual gesture shows significant differences in the /a-i/-context, where the greater quantity of a word is also accompanied by a longer gesture. In fact, statistical analysis



**Figure 4:** Tongue transition gesture. The kinematic characteristics are estimated for the duration, displacement (upper panels), and the peak velocity (bottom) of the articulatory movements. The panels show the stimuli with the CL Effect (left), VL Effect (mid), and VCL Effect (right).

reveals a three-way VCL Effect for the /t/-stimuli (at  $p < 0.05$  for  $\text{tapi}—\text{ta:p:i}$  and  $p < 0.001$  for  $\text{ta:p:i}—\text{ta:p::i}$  and  $\text{tapi}—\text{ta:p::i}$ ) and a significant two-way effect of Q1 and Q2 opposed to Q3 for /p/-initial stimuli ( $p < 0.05$  for  $\text{pa:p:i}—\text{pa:p::i}$  and  $p < 0.01$  for  $\text{papi}—\text{pa:p::i}$ ) reflecting the CL Effect.

Considering Vocalic Context Effect, the duration of the lingual transition is generally longer with movement from /a/ to /i/ compared to the transition from /i/ to /a/. The differences are statistically significant for  $\text{pa::pi}—\text{pi::pa}$ ,  $\text{pa:p:i}—\text{pi:p:a}$ ;  $\text{ta::pi}—\text{ti::pa}$  ( $p < 0.001$ ),  $\text{ta:p:i}—\text{ti:p:a}$  ( $p < 0.01$ ),  $\text{ta:p::i}—\text{ti:p::a}$  ( $p < 0.001$ ). The reverse pattern is not statistically significant, except for  $\text{tipa}—\text{tapi}$  pair ( $p < 0.05$ ) with longer transition gesture for the former. Consonant Context Effect elicits statistically significant differences for the Q2 pairs  $\text{papi}—\text{tap:i}$  ( $p < 0.01$ ),  $\text{pi:pa}—\text{ti:pa}$  ( $p < 0.05$ ), and  $\text{pi:p:a}—\text{ti:p:a}$  ( $p < 0.01$ ), where the transition gesture is longer with /p/-context.

There is no overall effect of Repetition on the duration of the lingual transition, and the interactions also show a slight decrease in duration for Q3  $\text{pap::i}$  and an increase in  $\text{pi::pa}$  ( $p < 0.05$  for both).

#### 3.4.2. Displacement

The y-axis of **Figure 4** shows the displacement of the transition gesture between the vowels surrounding the bilabial consonant. Statistical analysis revealed the effect of consonant lengthening (upper left panel in **Figure 4**) for the  $\text{papi}—\text{pap:i}—\text{pap::i}$  stimuli, where significant differences are found for Q1—Q2 ( $p < 0.01$ ) and for Q1—Q3 pairs ( $p < 0.001$ ), with more extensive transition gesture for Q1 compared to others. VL Effect depends on the vocalic context: With /a-i/-context, the difference is revealed as the opposition of Q1 vs. Q2 and Q3 ( $p < 0.001$ , except at  $p < 0.01$  for  $\text{papi}—\text{pa:pi}$ ) and the transition is larger with greater quantities, while with /i-a/-context, the contrast lies in Q1 and Q2 vs. Q3 ( $p < 0.001$ ) and the gesture is smaller in Q3 stimuli. The only significant effect of VCL is found for  $\text{tapi}—\text{ta:p:i}—\text{ta:p::i}$  stimuli with greater transition displacement in Q2 and Q3 vs. Q1 stimuli ( $p < 0.001$ ).

Considering Vocalic Context Effect, the transition gesture is more extensive with /i-a/-context than /a-i/-context in the case of  $\text{tipa}—\text{tapi}$  ( $p < 0.01$ ) and  $\text{pip::a}—\text{pap::i}$  ( $p < 0.05$ ),  $\text{tip::a}—\text{tap::i}$  ( $p < 0.01$ ), but with stimuli with an overlong vowel,  $\text{pi::pa}—\text{pa::pi}$  and  $\text{ti::pa}—\text{ta::pi}$  ( $p < 0.001$ ), the vocalic transition is significantly greater in the /a-i/ compared to the /i-a/-context. As for the stimuli that differ only in the stimulus initial consonant, the displacement is bigger in the /p/-initial words than with the /t/-initial words, but the difference is significant only for  $\text{papi}—\text{tapi}$  pair ( $p < 0.001$ ).

Repetition has a shortening main effect ( $p < 0.001$ ) on the displacement of the tongue transition gesture. The significant interactions between the repetition and the item appear for  $\text{tapi}$ ,  $\text{pipa}$  ( $p < 0.01$ ),  $\text{pa:pi}$ ,  $\text{pap:i}$  with an increasing effect on the displacement, and for  $\text{ta:pi}$ ,  $\text{tap:i}$ ,  $\text{tip:a}$  ( $p < 0.05$ ) with a decreasing effect.

#### 3.4.3. Peak velocity

The peak velocity of the lingual gesture is influenced by the CL Effect (see the left panel at the bottom in **Figure 4**), as increasing consonant quantity consistently slows down the transition movement. The model estimates show a three-way effect of CL for  $\text{papi}—\text{pap:i}—\text{pap::i}$  stimuli ( $p < 0.01$  for Q1—Q2 and Q2—Q3 comparisons,  $p < 0.001$  for Q1—Q3), and a two-way effect of Q1 and Q2 vs. Q3 for all the other triplets ( $p < 0.01$ , except  $p < 0.05$  for  $\text{tipa}—\text{tip::a}$ ).

There is almost no influence of *vowel lengthening* on the peak velocity of the lingual gesture; the only significant difference is elicited between  $\text{tapi}—\text{ta:pi}$  ( $p < 0.05$ ). Lengthening both vowel and consonant (see the rightmost plot at the bottom in **Figure 4**) reflects

the influence of consonantal quantity, as increase in word quantity is accompanied with slower gestures. These differences are significant for stimuli with /a-i/ vocalic context also forming a three-way distinction for papi-stimuli ( $p < 0.05$  for papi—pa:p:i,  $p < 0.01$  for pa:p:i—pa:p:i and papi—pa:p:i) and a difference between Q2 and Q3 ta:p:i—ta:p:i ( $p < 0.05$ ).

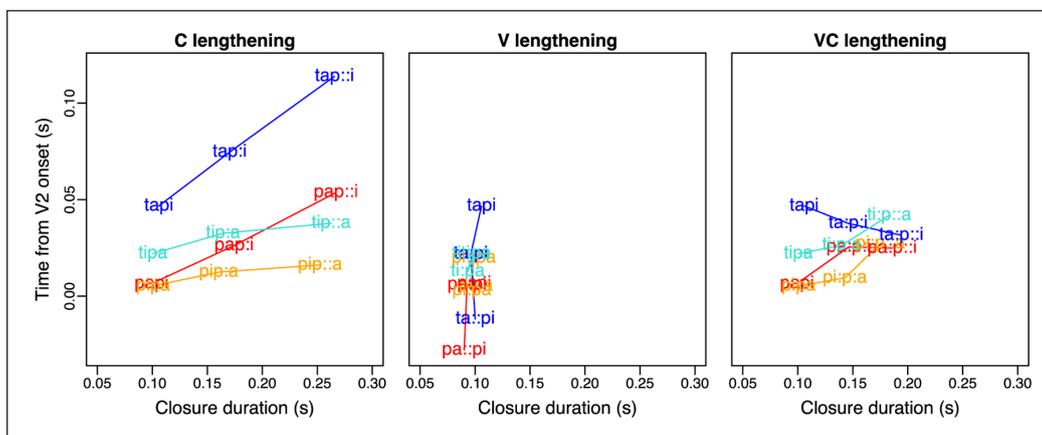
Regarding contextual effects, the velocity of the transition gesture is faster with /p/-initial stimuli in papi—tapi ( $p < 0.001$ ), pa::pi—ta::pi, pa:p:i—ta:p:i, and pa:p:i—ta:p:i ( $p < 0.01$ ). The repetition of a stimulus has a small decreasing effect ( $p < 0.05$ ) on the peak velocity and the interactions between an item and repetition show faster movements with repetition for pa:p:i ( $p < 0.05$ ) and pa:p:i ( $p < 0.01$ ), and slower gestures for pi::pa ( $p < 0.05$ ) and pi:p:a ( $p < 0.01$ ).

### 3.5. Lip-tongue coordination

**Figure 5** depicts the results for the coordination between lips and tongue, i.e., the temporal interval from the onset of the lip closing gesture for the bilabial consonant to the onset of the lingual gesture for the following vowel.

The interval between the onsets of the gestures is influenced by both consonantal and vocalic quantity in the /a-i/ vocalic context. As can be seen from **Figure 5**, the duration of the interval (y-axis) is increasing with growing consonantal length and decreasing with vowel length. The differences induced by CL Effect are significant for the Q1—Q3 pair papi—pap:i ( $p < 0.001$ ) and for /t/-stimuli tapi—tap:i—tap:i forming a three-way distinction ( $p < 0.01$  for Q1—Q2,  $p < 0.001$  for Q2—Q3 and Q1—Q3 comparisons). A shortening effect of VL on the interval duration shows a two-way opposition of Q1 and Q2 vs. Q3 for both triplets with /a-i/ vocalic context ( $p < 0.01$  for pa:p:i—pa::pi, papi—pa::pi, ta:pi—ta::pi,  $p < 0.001$  for tapi—ta::pi). In the case of the stimuli contrasting in both CL and VL (the rightmost panel in **Figure 5**), the temporal interval between lip-tongue gestures does not differ among the triplets.

As for the Vocalic Context Effect, the interval is generally longer in /a-i/ than in /i-a/-stimuli in words with consonant lengthening ( $p < 0.001$  for pap:i—pip:a,  $p < 0.05$  for tapi—tipa,  $p < 0.001$  for tap:i—tip:a and tap:i—tip:a), and the reverse pattern emerges in stimuli with Q3 vowel ( $p < 0.001$  for pa::pi—pi::pa and ta::pi—ti::pa). Consonant Context Effect shows a linear trend: With the /t/-context, the temporal interval between the gestures is consistently longer than with the /p/-context. Nevertheless, the differences are significant for the /a-i/ vocalic context for stimuli differing in consonant length, tapi—papi, tap:i—pap:i, and tap:i—pap:i ( $p < 0.001$ ).



**Figure 5:** Closing gesture displacement vs. the duration of C2-onset-to-V2-onset interval. The panels show the stimuli with the CL Effect (left), VL Effect (mid), and VCL Effect (right).

There is no overall effect of the repetition of a stimulus on the interval duration between the lip-tongue gestures. The only significant, but minute differences emerge for *pap:i* ( $p < 0.05$ ) where the gesture gets faster with the repetition and for *pi:pa* ( $p < 0.05$ ) with a slowing effect.

#### 4. Discussion

The results suggest that kinematic characteristics of the articulatory gestures are influenced by the quantity on the segment level showing different patterns depending on vocalic and consonantal context (in line with Šimko et al., 2014).

**Table 2** presents a summary of statistically significant effects of lengthening the intervocalic consonant (CL), the preceding vowel (VL) or both (VCL) on three kinematic characteristics (duration, displacement, peak velocity) of the gestures of lip closing, lip opening, and tongue transition as well as on the measure of the lip-tongue coordination.

For the lip closing gesture, quantity effects are realized in a consistent and expected way—the bigger the quantity of an intervocalic stop, the longer the movements (indicated by  $<$  in **Table 2**). Consonantal lengthening (see CL rows in **Table 2**) elicits a three-way effect with /i-a/ vocalic context and a two-way effect with /a-i/-context. The

**Table 2:** An overview of comparisons among estimates for different quantity realizations (Q1, Q2, Q3) for the kinematic characteristics of investigated gestures as well as the tongue-lips temporal coordination measure. The relational symbols ( $<$ ,  $>$ ) indicate a statistical significance of the given differences between estimates; equality sign (=) symbolizes no significant difference.

Gesture	Characteristic	Effect	papi	pipa	tapi	tipa
Lip closing	Duration	CL	Q1 < Q2 = Q3	Q1 < Q2 < Q3	Q1 = Q2 < Q3	Q1 < Q2 < Q3
		VL	Q1 < Q2 = Q3	Q1 < Q2 = Q3	–	Q1 < Q3
		VCL	Q1 < Q2 = Q3	Q1 < Q2 < Q3	Q1 < Q2 = Q3	Q1 < Q2 < Q3
	Displacement	CL	–	–	Q1 = Q2 < Q3	Q1 = Q2 < Q3
		VL	–	Q1 < Q2 = Q3	–	–
		VCL	Q1 < Q2 = Q3	Q1 < Q2 = Q3	Q1 < Q2	Q1 < Q3
	Peak velocity	CL	–	–	Q1 < Q3	Q2 < Q3
		VL	–	–	–	–
		VCL	–	–	–	–
Lip opening	Duration	CL	–	–	–	–
		VL	Q1 = Q2 > Q3	Q1 = Q2 > Q3	–	–
		VCL	Q2 > Q3	–	–	–
	Displacement	CL	–	–	–	–
		VL	Q1 = Q2 > Q3	Q1 = Q2 > Q3	–	–
		VCL	Q2 > Q3	–	–	–
	Peak velocity	CL	–	–	–	Q1 = Q2 < Q3
		VL	Q1 > Q3	–	–	–
		VCL	–	–	Q1 < Q2	–
Tongue transition	Duration	CL	Q1 = Q2 < Q3	–	Q1 = Q2 < Q3	Q2 < Q3
		VL	–	Q2 > Q3	Q1 < Q2 = Q3	Q1 = Q2 > Q3
		VCL	Q1 = Q2 < Q3	–	Q1 < Q2 < Q3	–
	Displacement	CL	Q1 > Q2 = Q3	–	–	–
		VL	Q1 < Q2 = Q3	Q1 = Q2 > Q3	Q1 < Q2 = Q3	Q1 = Q2 > Q3
		VCL	–	–	Q1 < Q2 = Q3	–
	Peak velocity	CL	Q1 > Q2 > Q3	Q1 = Q2 > Q3	Q1 = Q2 > Q3	Q1 = Q2 > Q3
		VL	–	–	Q1 < Q2	–
		VCL	Q1 > Q2 > Q3	–	Q2 > Q3	–
Lip-tongue coordination	Interval duration	CL	Q1 < Q3	–	Q1 < Q2 < Q3	–
		VL	Q1 = Q2 > Q3	–	Q1 = Q2 > Q3	–
		VCL	–	–	–	–

same tendency—longer movements with bigger quantity—applies in terms of vocalic lengthening (VL rows) which is marked by a two-way effect of Q1 vs. Q2 and Q3 in /p/-stimuli. Combining CL and VL Effects (VCL rows in **Table 2**) reflects the CL Effect and results in either a three-way or a two-way contrast of Q1 vs. Q2 and Q3. A similar two-way pattern influenced by quantity is shown for the displacement of the gesture. The spatial extent is greater in Q3 words compared to Q1 and Q2 words induced by CL Effect in /t/-initial stimuli and greater in Q2 and Q3 vs. Q1 stimuli elicited by VCL Effect in /p/-initial stimuli. The peak velocity of the lip closing gesture is generally not affected by the quantity; only some of the /t/-stimuli show faster gestures with greater quantity.

As the duration of the acoustic closure of the bilabial increases with consonantal length, the articulatory gestures associated with lip closure can be of longer duration (the compression of the soft lip tissue can continue longer after the occlusion is achieved) and, consequently, of greater extent. Our results are in line with the findings for Japanese, Tarifit Berber, Tashlhiyt Berber, and Moroccan Arabic (Löfqvist, 2006, 2007; Bouarourou et al., 2008; Ridouane, 2007; Zeroual et al., 2008), where geminates are reported to have longer durations of lip closing compared to singletons. Older articulatory studies on Estonian gemination also show an increasing contact area of the articulatory movements with longer consonants (Eek, 1970a, 1970b, 1970c, 1971a, 1971b, 1971c). It should be noted that these studies are based on material containing invariable segmental context. Similar direction for the displacement has been found in other languages as well, for example, the spatial extent of the lip closing gesture is larger for geminate bilabials than for singletons in Finnish, Japanese, and Italian (O'Dell et al., 2011; Löfqvist, 2005; Fivela & Zmarich, 2005).

On the contrary, quantity effects on the lip opening gesture appear to be less evident. We can see a significant shortening effect of vocalic quantity (VL rows in the **Table 2**) on the duration and displacement of the opening gesture in the /p/-context (shorter and smaller opening with Q1 and Q2 vowels compared to Q3 ones). There is no statistically significant difference with the /t/-context, which might be explained by the relatively small influence of /t/ on the neighboring segments, while another bilabial in the beginning of the word can cause smaller lip opening during the following vowels. The difference in the direction of the two consonantal gestures—lip closing and opening—might be due to compensatory durational alternations of V2 in Estonian where V2 is shortened when V1 is long and vice versa. Since the lip closing gesture already starts before the acoustical closure of the bilabial, the longer the V1, the more time there is to close the lips. In the case of the lip opening, the gesture is moving towards the following vowel, which is short when the V1 is long and, consequently, leaving less time for the opening gesture.

The analysis of the results revealed no clear patterns in the significant differences regarding the effect of Vocalic Context Effect (/a-i/ vs. /i-a/) on the kinematic characteristics of both the lip closing and lip opening gestures. Consonant Context Effect (/p/ vs. /t/) shows a general pattern of longer and larger consonantal gestures with /t/ as a stimulus initial consonant compared to the /p/-stimuli. The differences are mainly significant for Q3 pairs and, in particular, more consistently for the lip opening gesture duration. Differences between stimuli with a different word-initial consonant suggest the impact of gestural crowding (Beňuš & Šimko, 2014) as there is more time for the gestures in the /t/-context with only one bilabial closing movement compared to the /p/-stimuli with two closing gestures (i.e., a bilabial closing movement for the word-initial /p/ and another one for the word-medial /p/).

The lingual gesture of transition between the vowels surrounding the bilabial consonant is influenced by consonantal quantity as growing quantity degree is accompanied with longer transition gestures, indicating that during the longer bilabial closure there is more time for lingual movement. Consequently, the velocity of the inter-vocalic gesture

is slowed down during longer consonants. The opposition emerges generally as Q1 and Q2 vs. Q3. When the quantity distinction is realized by lengthening both the vowel and consonant, the effects on the kinematic characteristics of the lingual transition tend to follow the pattern induced by consonant lengthening for duration and peak velocity in stimuli with /a-i/ vocalic transition (temporally longer and slower gestures with increasing quantity). One of the anonymous reviewers proposed that the longer transitions could be due to avoiding the disruption of the timing of lip and tongue movements by changing the tongue movement trajectory, hence, while the lips are closed the tongue does not stop its movement, as according to our data, it slows down to keep the transition gesture longer. This interpretation is also in line with the coproduction theories suggesting that gestures for vowels and for consonants are produced separately but overlap in time (Fowler & Saltzman, 1993). Similar results—longer and slower transitions in geminate context—have also been found for Finnish (Šimko et al., 2014) and Japanese (Löfqvist, 2006).

The results for the displacement of the inter-vocalic gesture show significant differences elicited by the interaction of vocalic quantity (VL) and vocalic context. For the /a-i/-stimuli, the displacement is greater in Q2 and Q3 compared to Q1 words. In /i-a/-stimuli, the displacement is smaller with Q3 opposed to Q1 and Q2 words. Also, comparing the stimuli that differ only in their vocalic context (/a-i/ vs. /i-a/) shows a significantly greater displacement for the /a-i/-transition compared to /i-a/ in words with an overlong vowel and the reverse situation in words with an overlong consonant. The duration of the lingual transition tends to be longer for /a-i/-context than for /i-a/ in words with a long or overlong vowel or consonant. The velocity of the gesture is faster with /p/-initial stimuli compared to /t/-ones, emerging also in /a-i/-context. These patterns might be due to the fact that articulation of the vowel /a/ is less constrained and therefore more influenced by quantity (recall the results on articulatory variation for German lax/tense contrast by Hoole & Mooshamer, 2002, discussed in the Introduction) and, again, with the compensatory nature of the Estonian quantity system. As the tongue position is relatively constrained and less sensitive to the quantity for /i/ than for /a/, it is the length of the latter that impacts the kinematic characteristics more than that of the former. Therefore, in the /a-i/-context, the movement reflects the phonological length as captured by our variables (i.e., the length of the vowel V1 in our stimuli) while in the /i-a/-context, the effect is primarily due to the complementary length of V2.

The influence of both consonant lengthening and vowel lengthening on the coordination between the movements of the lips and the tongue is realized in the /a-i/ vocalic context. Increasing the quantity of an intervocalic consonant lengthens the interval duration between the two gestures, whereas /t/-initial stimuli show a ternary distinction. Since the bilabial closure is longer in bigger quantities, there is more time for the inter-vocalic transition to start later. Šimko et al., (2014) show similar results for Finnish singleton-geminate distinction, where a longer interval between the lip-tongue movement was found for geminates compared to singletons. Increasing vowel quantity, however, shortens the temporal interval of coordination resulting in a two-way contrast of Q1 and Q2 vs. Q3. Consequently, combining both vowel and consonant lengthening effects neutralizes these influences.

The transition movement with respect to the lip closing gesture starts earlier when the movement is from an overlong /a/ to /i/ compared to the transition from an overlong /i/ to /a/. In the /t/-context in stimuli differing only in consonant length, the transition starts later for /a-i/ compared to /i-a/-context. A general tendency of an earlier start of the lingual movement in the case of the /a-i/-transition was found in some other studies (Löfqvist & Gracco, 1999; Šimko & Cummins, 2009, 2010).

As the analysis reveals, the Estonian three-way quantity contrast is articulatorily realized in a rather complex way. Overall, as expected for a system contrast primarily by segmental duration, robust quantity effects are manifested in gestural duration, as we can see a quantity based two-way or three-way distinction for the lip closing gesture duration, while there is no influence of quantity on peak velocity of the gesture. Thus, from the perspective of the virtual target hypothesis suggested by Löfqvist (2005)—longer consonants have a higher virtual target for the lip movements—our results also show longer (and more extensive) movements with geminates, but with no changes in the speed of the articulators, which has also been found for Japanese labial stops and fricatives (Löfqvist, 2005).

The articulatory patterns shown in the study generally coincide with phonological descriptions of the Estonian quantity according to which a stressed Q1 syllable is considered short and light, Q2 long and light, and Q3 long and heavy. Shorter and smaller lip closing movements are found for words with short and light syllables compared to words with long stressed syllables, while the difference between the two types of long does not emerge in the /a-i/ context. Regarding the discussion of moraic representation of geminates (Q3 coda consonants can be linked to a mora and Q2 ones cannot), we would have expected longer and larger lip closing movements for Q3 compared to Q2, but the differences did not emerge in all contexts observed.

This distinction seems to be supported by the unstressed syllable. Phonologically the unstressed syllables in Estonian do not have distinctive length oppositions independent of the stressed syllables, but their duration lengthens or shortens in the opposite direction of the stressed syllable. We would observe this pattern for the lip opening gesture duration and displacement, with shorter and smaller movements in Q3 compared to Q2 feet, albeit only in the case of vocalic lengthening in the /p/-stimuli. The shorter and smaller lip opening gesture in Q3 could be explained by the suggestion of Prillop (2015) who accounts the shortening and reduction of the unstressed syllables of Q3 feet for not belonging to the foot because the Q3 stressed syllable constitutes a foot on its own. For Q1 and Q2 the feet are disyllabic. The articulatory pattern for the lip opening gesture also lends support to the idea of weak and strong moras of Prillop (2013) discussed in the Introduction. Strong moras have phonetically been described with longer or more intense sounds, and articulatory results seem to confirm this. Whereas a strong mora causes final lengthening at the end of Q1 foot and Q3 syllable, the lip opening movements are longer and of greater extent in Q1 and Q2 feet with a strong mora in the second syllable, but shorter and smaller for Q3. In addition, articulatory level may contribute to the perception, as in previous studies it has been shown that the unstressed syllables have a role in distinguishing Q2 from Q3 (Eek & Meister, 2003).

Regarding the other two characteristics—tongue transition and lip-tongue coordination—the articulatory patterns show an influence of consonantal quantity (CL Effect). It tends to have a uniform effect on the duration and peak velocity of the tongue transition gesture with longer and slower transitions with greater quantity of an intervocalic consonant. The timing between the onsets of the lip closing gesture and the lingual gesture (lip-tongue coordination in **Table 2**) is also delayed in the case of a higher quantity of an intervocalic consonant. These results are in line with Smith's (1995) "combined vowel-and-consonant timing" model where the timing of the lingual transition is also affected by the increasing consonant length. Smith pointed out that this is a characteristic of a mora-timed language Japanese. As discussed in the introduction of our paper, the Estonian quantity system cannot easily be fitted into moraic theory. In spite of the fact that the articulatory behavior of the gestures studied here can in a few aspects be accounted for by the moraic point of view, our results mostly lend support to the phonological approach of the manifestation of

a three-way quantity distinction in a disyllabic foot, as, for example, we saw an influence of between-segment durational alternations on the lip closing and opening gestures.

Our data, however, do not reveal any clear articulatory ‘signature’ that would reliably correlate with the complex system of Estonian quantity contrast in all phonemic contexts and across all speakers. While duration of the lip closing gesture, for example, exhibits robust sensitivity to quantity patterns in the /i-a/-context, Q2 is not clearly separated from Q1 or Q3 by this measure in /a-i/-words. Importantly, this picture does not arise purely as a result of the very conservative nature of statistical modeling used in this paper; other, somewhat more restrictive modeling approaches (e.g., mixed-effect models with interacting quantity, consonantal, and vocalic context; not reported here), yield results that differ only in small details. Situation is thus somewhat parallel to the strong phonemic context sensitivity of acoustic durational patterns reported for the same material by Lippus and Šimko (2015): Parameterization of the kinematic characteristics of articulation underlying the Estonian quantity does not simply reflect phonologically conceptualized quantity patterns but arises as a solution of the task of sufficiently distinguishing between the contrastive quantity patterns for the particular set of gestures involved.

The lack of clear patterns might, of course, be simply due to variations between speakers; speaker specific characteristics will thus be a natural task for subsequent work. Also, in the material there are both meaningful and nonce words. An experimental material with only real Estonian words—however challenging its design might be—may reveal regularities that are not manifested for nonsense expressions.

Finally, this study highlights the necessity to study complex phonological systems such as Estonian quantity on a varied speech material that accounts for articulatory variation among segments. If studied in isolation, each individual segmental context (columns in **Table 2**) would present a different picture about the manifestation of the three-way distinction among articulatory characteristics. As we saw in the discussion, most of these differences can be attributed to the articulatory characteristics of individual segments and their mutual interactions, such as variability, intrinsic duration, movement direction, and gestural crowding. In order to get a more comprehensive picture of the articulatory realization of the Estonian quantity system and avoid a possibility of drawing conclusions that reflect properties of individual segments rather than effects of quantity itself, we need to combine the articulatory correlates of quantity marking that are manifested in different words containing varied segmental material.

## 5. Conclusions

In this paper, we studied the articulatory realization of the complex phonological quantity contrast in Estonian. The results show an influence of quantity on the kinematic characteristics (gesture duration, spatial extent, peak velocity) and on intergestural coordination. In several cases, the influence is, however, manifested in a non-linear way through interplays with the quality of segments. In fact, the analysis revealed a strong context sensitivity both with respect to the vocalic context and the preceding consonant, and as a consequence, on the articulatory level we can see a context dependent quantity contrast.

## Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Mixed effect model outputs for articulatory measures. DOI: <https://doi.org/10.5334/labphon.117.s1>
- **Appendix B.** The post-hoc pairwise comparisons. DOI: <https://doi.org/10.5334/labphon.117.s2>

## Acknowledgements

We would like to thank all test subjects and we are also very grateful to two anonymous reviewers for their useful comments on earlier versions of the paper. We are thankful to Mona Lehtinen for her assistance during the recordings and Jari Lipsanen for his advice on the statistical design of the study. We would like to thank Eva Liina Asu for proof-reading the paper. This research was supported by the Estonian Research Agency grant No. IUT2-37.

## Competing Interests

The authors have no competing interests to declare.

## References

- Baayen, R. H., & Milin, P. 2010. Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28. DOI: <https://doi.org/10.21500/20112084.807>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Beňuš, Š. 2011. Control of phonemic length contrast and speech rate in vocalic and consonantal syllable nuclei. *The Journal of the Acoustical Society of America*, 130(4), 2116–2127. DOI: <https://doi.org/10.1121/1.3624824>
- Beňuš, Š., & Šimko, J. 2014. Emergence of prosodic boundary: Continuous effects of temporal affordance on inter-gestural timing. *Journal of Phonetics*, 44, 110–129. DOI: <https://doi.org/10.1016/j.wocn.2013.12.005>
- Bouarourou, F., Vaxelaire, B., Ridouane, R., Hirsch, F., & Sock, R. 2008. Gemination in Tarifit Berber: X-ray and acoustic data. In: Sock, R., Fuchs, S., & Laprie, Y. (eds.), *Proceedings of the 8th International Seminar on Speech Production*, 117–120. Strasbourg, France. Retrieved from: <http://issp2008.loria.fr/Proceedings/PDF/issp2008-23.pdf>.
- Bye, P. 1997. A generative perspective on ‘overlength’ in Estonian and Saami. *Estonian Prosody: Papers from a Symposium. Proceedings of the International Symposium on Estonian Prosody*, 36–70. Tallinn, Estonia, October 29–30, 1996, Tallinn.
- Eek, A. 1970a. Articulation of the Estonian sonorant consonants. II. [r]. *Eesti NSV Teaduste Akadeemia Toimetised. Ühiskonnateaduste Seeria*, 19(3), 296–310.
- Eek, A. 1970b. Articulation of the Estonian sonorant consonants. I. [n] and [l]. *Eesti NSV Teaduste Akadeemia Toimetised. Ühiskonnateaduste Seeria*, 19(1), 103–121.
- Eek, A. 1970c. Some coarticulation effects in Estonian. *Soviet Fenno-Ugric Studies*, 6(2), 81–85.
- Eek, A. 1971a. Articulation of the Estonian sonorant consonants. III. Palatalized [n] and [l]. *Eesti NSV Teaduste Akadeemia Toimetised. Ühiskonnateaduste Seeria*, 20(2), 173–191.
- Eek, A. 1971b. Articulation of the Estonian sonorant consonants. IV. [m]. *Soviet Fenno-Ugric Studies*, 7(3), 161–168.
- Eek, A. 1971c. Articulation of the Estonian sonorant consonants. V. [n]. *Soviet Fenno-Ugric Studies*, 7(4), 259–268.
- Eek, A. 1974. Observations on the duration of some word structures: I. In: Eek, A. (ed.), *EPP 1974. To Honour Academician Paul Ariste on the Occasion of his Seventieth Birthday*, 18–32. February 3, 1975. Tallinn: Academy of Sciences of the Estonian S.S.R.
- Eek, A. 2008. *Eesti keele foneetika*. Tallinn: Tallinna Tehnikaülikooli Kirjastus.
- Eek, A., & Meister, E. 1997. Simple perception experiments on Estonian word prosody: Foot structure vs. segmental quantity. In: Lehiste, I., & Ross, J. (eds.), *Estonian prosody: Papers from a symposium*, 71–99. Tallin, Estonia: Institute of Estonian Language.

- Eek, A., & Meister, E. 1998. Quality of standard Estonian vowels in stressed and unstressed syllables of the feet in three distinctive quantity degrees. *Linguistica Uralica*, 34(3), 226–233.
- Eek, A., & Meister, E. 2003. Domain of the Estonian Quantity Degrees. Evidence from Words Containing Diphthongs. In: Solé, M.-J., Recasens, D., & Romero, J. (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*, 2039–2042.
- Eek, A., & Meister, E. 2004. Foneetilisi katseid ja arutlusi kvantiteedi alalt (II). Takt, silp ja vâlde. *Keel ja Kirjandus*, 4–5, 251–271, 336–357.
- Fowler, C. A. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 8, 113–133. Retrieved from: <http://www.ling.uni-potsdam.de/~gafos/courses/LABEX-Paris2016-Lecture1/Fowler1980JPhonExtrinsicIntrinsic.pdf>.
- Fowler, C. A., & Saltzman, E. 1993. Coordination and coarticulation in speech production. *Language and Speech*, 36(2–3), 171–195. Retrieved from: [http://www.linguistics.berkeley.edu/~kjohnson/LSA317/lsa317\\_fowler\\_saltzman1993.pdf](http://www.linguistics.berkeley.edu/~kjohnson/LSA317/lsa317_fowler_saltzman1993.pdf).
- Gili Fivela, B., & Zmarich, C. 2005. Italian Geminate under Speech Rate and Focalization Changes: Kinematic, Acoustic, and Perception Data. In: *Proceedings of Interspeech 2005*, 2897–2900. Lisbon, Portugal.
- Gili Fivela, B., Zmarich, C., Perrier, P., Savariaux, C., Tisato, G., & others. 2007. Acoustic and kinematic correlates of phonological length contrast in Italian consonants. In: *Proceedings of the XVIth International Congress of Phonetic Sciences*. Saarbrücken, Germany.
- Harrington, J., Fletcher, J., & Roberts, C. 1995. Coarticulation and the accented/unaccented distinction: evidence from jaw movement data. *Journal of Phonetics*, 23(3), 305–322. DOI: [https://doi.org/10.1016/S0095-4470\(95\)80163-4](https://doi.org/10.1016/S0095-4470(95)80163-4)
- Hayes, B. 1995. *Metrical Stress Theory*. Chicago-London.
- Hertrich, I., & Ackermann, H. 1997. Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *Journal of the Acoustical Society of America*, 102, 523. DOI: <https://doi.org/10.1121/1.419725>
- Hint, M. 1973. *Eesti keele sõnafonoloogia I. Rõhusüsteemi fonoloogia ja morfofonoloogia põhiprobleemid*. Tallinn: Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut.
- Hoole, P., & Mooshammer, C. 2002. Articulatory analysis of the German vowel system. In: Auer, P., & Gilles, P. P. (eds.), *Silbenschnitt und Tonakzente*, 129–152. Niemeyer, Tübingen. DOI: <https://doi.org/10.1515/9783110916447.129>
- Hoole, P., & Zierdt, A. 2010. Five-dimensional articulography. In: Maassen, B., & van Lieshout, P. (eds.), *Speech Motor Control: New Developments in Basic and Applied Research*, 331–350. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199235797.003.0020>
- Kager, R. 1993. Alternatives to the Iambic-Trochaic Law. *Natural Language & Linguistic Theory*, 11, 381–432. DOI: <https://doi.org/10.1007/BF00993165>
- Kroos, C., Hoole, P., Kühnert, B., & Tillmann, H. G. 1997. Phonetic evidence for the phonological status of the tense-lax distinction in German. *Forschungsberichte-Institut Für Phonetik Und Sprachliche Kommunikation Der Universität München*, 35, 17–25.
- Lehiste, I. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lehiste, I. 1990. Phonetic investigation of metrical structure in orally produced poetry. *Journal of Phonetics*, 18, 123–133.
- Lehiste, I. 2003. Prosodic change in progress: From quantity language to accent language. In: Fikkert, P., & Jakobs, H. (eds.), *Development in prosodic systems*, 47–65. Berlin/New York: Mouton de Gruyter.

- Lehiste, I., Morton, K., & Tatham, M. A. A. 1973. An instrumental study of consonant gemination. *Journal of Phonetics*, 1, 131–148.
- Lehtonen, J. 1970. Aspects of quantity in standard Finnish. In: *Studia philologica Jyväskyläensia*, 6. Gummérus Jyväskylä.
- Liiv, G. 1961. On qualitative features of Estonian stressed monophthongs of three phonological degrees of length. *Eesti NSV Teaduste Akadeemia Toimetised. Ühiskonnateaduste Seeria*, 10(1–2), 41–66, 113–131.
- 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. DOI: <https://doi.org/10.1016/j.wocn.2012.09.005>
- Lippus, P., Pajusalu, K., & Allik, J. 2009. The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics*, 37(4), 388–396. DOI: <https://doi.org/10.1016/j.wocn.2009.07.002>
- Lippus, P., Pajusalu, K., & Allik, J. 2011. The role of pitch cue in the perception of the Estonian long quantity. In: Frota, S., Elordieta, G., & Prieto, P. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, 231–242. Dordrecht: Springer Netherlands. Retrieved from: [http://www.springerlink.com/index/10.1007/978-94-007-0137-3\\_10](http://www.springerlink.com/index/10.1007/978-94-007-0137-3_10).
- Lippus, P., & Šimko, J. 2015. Segmental context effects on temporal realization of Estonian quantity. In: Wolters, M., Livingstone, J., Beattie, B., Smith, R., MacMahon, M., Stuart-Smith, J., & Scobbie, J. (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5. Glasgow: University of Glasgow. Retrieved from: <http://www.icphs2015.info/pdfs/Papers/ICPHS0318.pdf>.
- Löfqvist, A. 2005. Lip kinematics in long and short stop and fricative consonants. *The Journal of the Acoustical Society of America*, 117(2), 858. DOI: <https://doi.org/10.1121/1.1840531>
- Löfqvist, A. 2006. Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese. *The Journal of the Acoustical Society of America*, 120(5 Pt 1), 2872. DOI: <https://doi.org/10.1121/1.2345832>
- Löfqvist, A. 2007. Tongue movement kinematics in long and short Japanese consonants. *The Journal of the Acoustical Society of America*, 122(1), 512. DOI: <https://doi.org/10.1121/1.2735102>
- Löfqvist, A., & Gracco, V. L. 1997. Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech, Language, and Hearing Research*, 40(4), 877–893. DOI: <https://doi.org/10.1044/jslhr.4004.877>
- Löfqvist, A., & Gracco, V. L. 1999. Interarticulator programming in VCV sequences: Lip and tongue movements. *Jasa*, 105, 1864–1876. DOI: <https://doi.org/10.1121/1.426723>
- Meister, E., & Werner, S. 2015. Comparing palatography patterns of Estonian consonants across time. In: Wolters, M., Livingstone, J., Beattie, B., Smith, R., MacMahon, M., Stuart-Smith, J., & Scobbie, J. (eds.), *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5. Glasgow: University of Glasgow. Retrieved from: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0652.pdf>.
- Meister, E., Werner, S., & Meister, L. 2011. Short vs. long category perception affected by vowel quality. In: Lee, W.-S., & Zee, E. (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, 1362–1365. Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong.
- O'Dell, M., Šimko, J., Nieminen, T., Vainio, M., & Lehtinen, M. 2011. Relative timing of bilabial gesture in Finnish. In: *Proceedings of the 17th International Congress of Phonetic Sciences*. Hongkong.

- Öhman, S. E. G. 1966. Coarticulation in VCV Utterances: Spectrographic Measurements. *Jasa*, 39(1), 151–168. DOI: <https://doi.org/10.1121/1.1909864>
- Ostry, D. J., & Munhall, K. G. 1985. Control of rate and duration of speech movements. *Jasa*, 77(2), 640–648. DOI: <https://doi.org/10.1121/1.391882>
- Prillop, K. 2013. Feet, syllables, moras and the Estonian quantity system. *Linguistica Uralica*, 49(1), 1 – 29. DOI: <https://doi.org/10.3176/lu.2013.1.01>
- Prillop, K. 2015. Hääliku-, silbi- ning jalavälde: ühe nähtuse mitu tahku. *Journal of Estonian and Finno-Ugric Linguistics*, 6(3), 169 – 195. DOI: <https://doi.org/10.12697/jeful.2015.6.3.07>
- Prince, A. S. 1980. Metrical Theory for Estonian quantity. *Linguistic Inquiry*, 11, 511–562.
- Ridouane, R. 2007. Gemination in Tashlhiyt Berber: an acoustic and articulatory study. *Journal of the International Phonetic Association*, 37(02), 119–142. DOI: <https://doi.org/10.1017/S0025100307002903>
- Šimko, J., & Cummins, F. 2009. Sequencing of Articulatory Gestures using Cost Optimization. In: *Proceedings of Interspeech 2009*, 60–63. Brighton, UK.
- Šimko, J., & Cummins, F. 2010. Embodied Task Dynamics. *Psychological Review*, 117(4), 1229–1246. DOI: <https://doi.org/10.1037/a0020490>
- Šimko, J., & Cummins, F. 2011. Sequencing and Optimization within an Embodied Task Dynamic Model. *Cognitive Science*, 35(3), 527–562. DOI: <https://doi.org/10.1111/j.1551-6709.2010.01159.x>
- Šimko, J., O'Dell, M., & Vainio, M. 2014. Emergent consonantal quantity contrast and context-dependence of gestural phasing. *Journal of Phonetics*, 44, 130–151. DOI: <https://doi.org/10.1016/j.wocn.2013.11.006>
- Smith, C. L. 1995. Prosodic patterns in the coordination of vowel and consonant gestures. *Papers in Laboratory Phonology IV, Phonology and Phonetic Evidence*. CUP, 205–222.
- Viitso, T.-R. 2003. Structure of the Estonian language: Phonology, morphology and word formation. In: Erelt, M. (ed.), *Estonian language. Linguistica Uralica supplementary series*, 1, 9–129. Tallinn: Estonian Academy Publishers.
- Zeroual, C., Hoole, P., & Gafos, A. I. 2008. Spatio-temporal and kinematic study of Moroccan Arabic coronal geminate plosives. In: Sock, R., Fuchs, S., & Laprie, Y. (eds.), *Proceedings of the 8th International Seminar on Speech Production*, 135–138. Strasbourg, France. Retrieved from: <http://issp2008.loria.fr/Proceedings/PDF/issp2008-27.pdf>.

**How to cite this article:** Türk, H., Lippus, P. and Šimko, J. 2017 Context-dependent articulation of consonant gemination in Estonian. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1): 26, pp. 1–26, DOI: <https://doi.org/10.5334/labphon.117>

**Submitted:** 19 September 2017    **Accepted:** 19 September 2017    **Published:** 03 November 2017

**Copyright:** © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 