# Peak Alignment in Estonian

Mareike Plüschke

München 2013

# Peak Alignment in Estonian

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Philosophie

an der Ludwig-Maximilians-Universität München

vorgelegt von

Mareike Plüschke

Erstgutachter: Prof. Dr. Jonathan Harrington
Zweitgutachter: PD Dr. phil. habil. Philip Hoole
Tag der mündlichen Prüfung: 25. Juni 2013

# Acknowledgements

First of all I would like to thank my supervisor Jonathan Harrington for all his support, suggestions, feedback, encouragement and help in the past years. I am especially grateful that you helped me focus on the important things.

I am also very grateful to my second supervisor Phil Hoole, and Elena Skribnik who is not only the third member of the examination committee but also supported me in finding a PhD scholarship by writing quite many references.

I thank all my colleagues from the IPS and members of the PhD & Post-Doc colloquium who gave me a lot of feedback and always helped me willingly whenever I needed it. In particular, I would like to thank Klaus Jänsch and Christoph Draxler for help with Speechrecorder and Wikispeech, which made it a pleasure to do the recordings. I am also grateful to Klaus for all his technical support. Many thanks go to Florian Schiel for making MAUS 'speak' Estonian and helping me with a tricky R-script. You saved me a lot of time. I am also grateful to Ulrich Reubold and Rafael Winkelmann who helped me out when I reached my personal limits with R. I would like to thank Olga Dioubina for a wonderful praat script. Thank you very much also to Sibylle von Tiedemann from the ESP PhD program.

There are several people from Estonia who helped me in some form or the other in the past years. I would like to thank all phoneticians from the Universities of Tartu and Tallinn and all my friends in Estonia who provided me with help. Especially I would like to thank (in alphabetical order) Eva Liina Asu, Pärtel Lippus, Nele Salveste, Pire Teras and Tuuli Tuisk from the University of Tartu and Einar and Lya Meister from the University of Tallinn. Thank you, Eva Liina, for feedback on my work and answering all the questions I had concering Estonian phonetics. Many thanks to you, Pärtel, for feedback and for being always a willing subject for my recordings. Nele, I am very grateful for all our discussions about Estonian phonetics and my speech material. I am also grateful that you got so many of your friends to participate in my recordings. Pire, your list with vowel and consonant quantity triplets is absolutely great and I am very thankful for all our conversations about my speech material. Thank you very much, Tuuli, for helping me with the Estonian language and for not only participating in my recordings but for also providing me with other subjects. Einar and Lya, it was a pleasure to use your facilities at the University of Tallinn and to record you and your colleagues. Furthermore I would like to thank all my Estonian subjects who participated in my recordings. Suur tänu - thank you very much to all of you.

I would like to thank Mariapaola D'Imperio and her colleagues, especially Francesco Cangemi and Pauline Welby, from the Laboratoire Parole & Langage in Aix-en-Provence for feedback and help in my early days of peak alignment research. I am grateful to Bob Ladd for discussions about my work when I just started with my research. I would also like to thank all other people who discussed my work with me during the past years.

My dissertation years did not only consist of work, but also of nice collegial times at the IPS. I had a lot of fun with my lunch mates and the 'coffee group' guys. Thanks to all of you.

I am very grateful to the Studienstiftung des Deutschen Volkes for providing me with a three year PhD fellowship. Thanks to the fellowship I could fully concentrate on my dissertation.

I would like to thank Saskia H. for encouraging me and helping me in both easy and difficult times during my PhD studies.

Last but not least I would like to thank my husband Andreas. Thank you for your unconditional love and support. It is impossible for me to put my gratitude to you in words here.

# Contents

# List of Tables

# List of Figures

# Summary

Earlier studies of Estonian showed that vowel quantity words (i.e. words differing only in vowel quantity) produced with an H*+L pitch accent differed in their peak alignment: While words with a short and a long vowel had a peak late in the vowel of the stressed syllable, words with an overlong vowel were characterised by a peak earlier in the vowel (e.g. Asu et al., 2009; Lippus et al., 2013). The main aim of this dissertation is to shed light on these peak alignment differences: firstly, whether these alignment differences can be explained with the help of a segmental anchor; secondly, whether alignment is similarily affected by quantity differences in consonants and vowels; thirdly, whether such alignment differences are stable in regard to the prosodic context, more precisely in regard to the number of post-focal unstressed syllables (i.e. the vicinity to the sentence boundary) and a variation of the speaking rate. Additionally, not only the peak alignment in regard to the vicinity of an upcoming sentence boundary was investigated, but also the influence of the sentence boundary on segment durations (phrase-final lengthening - PFL). Previous studies (e.g. Krull, 1997; Asu et al., 2009) showed that PFL occurs in Estonian, but it was not studied yet whether PFL affects vowel and consonant quantity words differently. Furthermore no attempt made to explain PFL in Estonian with the help of abstract phonological models. The purpose of this dissertation is to fill this gap.

This dissertation contains three different experiments which are presented in one chapter each. The first experiment (chapter 2) explored the influence of the upcoming sentence boundary and its interaction with vowel (VQ) and consonant (CQ) quantity on the peak alignment of falling nuclear H*+L pitch accents. Disyllabic target words ($C_1V_1C_2V_2$) only differing in either the quantity of $V_1$ (VQ-words) or $C_2$ (CQ-words) were embedded in two different carrier sentences: in one carrier sentence the target word was followed by two unstressed syllables (long tail context) and in the other by none (short tail

context). All target words occured in three quantity degrees: short (Q1), long (Q2) and overlong (Q3). There were two main results: (1) In the short tail context the peak was aligned earlier. (2) The peak alignment of VQ- and CQ-words was similar. Quantity degree differences of both VQ- and CQ-words were cued by the peak alignment in proportion to the $V_1C_2$-duration. The proportional peak alignment had the order Q3 $<$ Q2 $<$ Q1, where $<$ denotes that the peak of Q3-words was proportionally timed earlier than the peak of Q2-words and so on.

The second experiment (chapter 3) analysed the influence of the sentence boundary, i.e. phrase-final lengthening (PFL), on the segment durations of VQ- and CQ-words. The data used for the analysis was the same as in the first experiment. There were two main results: (1) The domain of PFL in Estonian was the main bearer of the quantity contrast, i.e. $V_1$ in VQ-words and $C_2$ in CQ-words and can be best accounted for in terms of a Structure-based model for explaining PFL (Turk and Shattuck-Hufnagel, 2007). (2) Progressive lengthening, i.e. the nearer a segment is to the final boundary the more it is lengthened, occured in the data if the lengthened segments were not in adjacent word-final position.

The third experiment (chapter 4) investigated whether speaking rate influences the alignment of the peak. VQ- and CQ-words were embedded in carrier sentences with one unstressed syllable following the target word. They were read in normal and fast speaking rate. There were two main results: (1) In both VQ- and CQ-words the peak alignment in proportion to the $V_1C_2$-duration had the order Q3 $<$ Q2 $<$ Q1, where $<$ denotes that the peak of Q3-words was aligned earlier than the peak of Q2-words and so on. (2) Speaking rate did not influence the peak alignment in proportion to the $V_1C_2$-duration.

The results of this dissertation favour in interpretation in the sense of the segmental anchoring hypothesis (see e.g. Ladd et al., 1999, 2000; Schepman et al., 2006) that tonal targets are anchored with specific points of the segmental string. The results of the current dissertation created the impression that in Estonian the offset of the first mora could be the anchorpoint for the peak - regardless of quantity degree and type. Differences in the proportional peak alignment emerge because the anchorpoint interacts with the temporal correlates of the quantity contrast. Compatibly with Ladd (2008), the results of the dissertation also show that peak alignment in Estonian is influenced by phonologically-induced (an increase in the number of post-focal syllables) but not phonetically-induced (faster speaking rate) time pressure.

# Chapter 1

# Introduction

The alignment of tonal targets with the segmental string and especially its stability in regard to prosodic factors (e.g. time pressure, syllable structure, phonological vowel length) is a currently much debated topic, and it was shown to be language-dependent (e.g. Atterer and Ladd, 2004; Schepman et al., 2006). The purpose of this dissertation is to investigate the influence of phonological quantity on the large unresolved issue of how tonal alignment interacts with other prosodic factors. Estonian is ideally suited for such an investigation because it has a three-way quantity contrast for both vowels and consonants.

The first subsection of this chapter contains an overview of the Estonian language. The second subsection is concerned with intonation, particularly with intonation models in general, a model of Estonian intonation specifically and an introduction into tonal alignment research.

## 1.1 The Estonian Language

Estonian is a Finno-Ugric language spoken by about one million people in Estonia and about 150.000 people outside Estonia (Asu and Teras, 2009), living mainly in Australia, Canada, Latvia, Russia, Sweden and the USA (Viitso, 1998). Together with Finnish, Ingrian, Karelian, Livonian, Veps and Votic, Estonian belongs to the Finnic branch of the Finno-Ugric languages (Viitso, 2003a).

Estonian is a dialect rich language with three main dialect groups (North-Eastern Coastal Estonian, North Estonian and South Estonian), which can be further divided into 8-10 dialects and 105-120 subdialects (Pajusalu, 2003). Nevertheless, since the beginning of the 20th century Standard Estonian, which is investigated in this dissertation, gained more and more influence over the dialects and is nowadays the language of all educated Estonians (Laanekask and Erelt, 2003). Standard Estonian is a compromise between various dialects, the influence of foreign languages and a language reform. Nevertheless, Standard Estonian has more common features with the North Estonian dialects than the South Estonian dialects (Pajusalu, 2003).

The Estonian sound system consists of 17 consonant phonemes /p t tʲ k m n nʲ r f v s sʲ ʃ h l lʲ j] and nine vowel phonemes [i y e ø æ u ɤ o ɑ/. All nine vowel phonemes can occur in primary stressed (initial) syllables or as the first component of a diphthong, but only the vowels /ɑ e i o u/ can occur in non-initial syllables or as the second component of a diphthong (Asu and Teras, 2009). In total 36 different diphthongs are used in Estonian (Asu and Teras, 2009).

Estonian is a quantity language with a three-way quantity contrast (short, long and overlong) for both vowels and consonants. All vowel and consonant phonemes can particpate in the quantity distinction, i.e. they can occur as short (Q1), long (Q2) and overlong (Q3) in the first, stressed syllable of a disyllabic word (Ross and Lehiste, 2001). It is not the case that long sounds are always two times longer and overlong sounds three times longer than their short counterparts. In fact long vowels were found to be about 1.6 - 2.0 times longer and overlong vowels about 2.1 - 2.9 times longer than the appropriate short vowels (duration data from various studies: Asu et al., 2009; Eek, 1974; Eek and Meister, 1998, 2003; Krull, 1997; Lehiste, 1960; Liiv, 1961; Meister and Meister, 2011; see Meister et al., 2011 for an overview). Consonant durations vary more depending on the consonant. In the studies of Eek (1974) Lehiste (1966), and Plüschke (2009) long consonants were 1.31 - 3.45 times longer and overlong consonants 1.83 - 7,03 times longer than their short counterparts.

Vowel quality in Estonian varies to some extent with the quantity degrees and stress. In spontanous speech short stressed vowels tend to be more centralised than their long and overlong counterparts and deaccented stressed vowels more centralised than their accented matches (Lippus, 2010; Lippus et al., 2013). In read speech the difference in vowel quality between the three quantity degrees was found to be smaller than one Bark (Eek and Meister, 1998). Vowels in unstressed syllables were reduced to some degree in both read (Eek and Meister, 1998) and spontanous speech (Lippus, 2010;

Lippus et al., 2013). While in read speech the reduction of unstressed vowels was greatest in Q3-words and least in Q1-words (Eek and Meister, 1998), in spontanous speech no systematic correlation between the reduction of unstressed vowels and the quantity degrees or senctence accentuation was found (Lippus, 2010; Lippus et al., 2013).

In native Estonian words, primary stress is fixed and usually falls on the first syllable. Exceptions are some interjections such as e.g. *aitäh* ('thanks') or *ennäe* ('behold'), where the primary stress is not on the first syllable. In borrowings, primary stress can also fall on non-initial syllables (Viitso, 2003b). The maximum stress foot in native Estonian words is trisyllabic, consisting of the first, primary stressed syllable and one (disyllabic foot) or two (trisyllabic foot) unstressed syllables. Words containing more than three syllables are formed combinations of monosyllabic, disyllabic and trisyllabic feet (Asu and Teras, 2009). In tetrasyllabic words the third syllable is secondary-stressed. In longer words, secondary stress concerns odd-numbered syllables following the primary stressed syllable. Exceptions are suffixes such as *-mine* (*laulmine* [ˈlauːlˌmine] 'singing') which always get stressed (Asu and Teras, 2009).

Rhythmically Eek and Help (1987) classified Estonian as syllable-timed, but newer studies showed that Estonian is neither purely syllable- nor stress-timed, but instead exhibits features of both groups (Asu and Nolan, 2006; Grabe and Low, 2002; Nolan and Asu, 2009).

The basic quantity distinction in Estonian is between short (Q1), long (Q2) and overlong (Q3). The quantity distinction is not the property of a single vowel or consonant, but operates on the level of a disyllabic metric foot, of which the first syllable is always stressed and the second syllable is unstressed (Lehiste, 1960). On the phonemic level there is only a two-way (short vs. long) quantity contrast (Eek and Meister, 1997, 2003; Lehiste, 1997a, 2003) The quantity distinction affects either the vowel of the first syllable (VQ - vowel quantity) or the intervocalic consonant (CQ - consonant quantity) of the disyllabic foot ($C_1V_1C_2V_2$). In unstressed syllables, there is no quantity distinction, but Estonian has a tendency towards foot isochrony which is reflected in the syllable durations: the longer the first syllable, the shorter the second one, i.e. the second syllable is longest in Q1-feet and shortest in Q3-feet (e.g. Krull, 1999; Lehiste, 1997a, 2003; Lippus et al., 2013; Nolan and Asu, 2009). An example for a VQ-triplet is VQ1 *sada* [sɑtɑˑ][1]

[1]Due to the foot isochrony the vowel of the second syllable is half-long in Q1-words.

'hundred' (nominative singular), VQ2 *saada* [sɑːtɑ] 'send' (imperative, 2nd singular) and VQ3 *saada* [sɑːːtɑ] 'to get' (infinitive). CQ1 *kala* [kɑlɑˑ][2] 'fish' (nominative singular), CQ2 *kalla* [kɑllɑ] 'calla lily' (nominative singular) and CQ3 *kalla* [kɑlːlɑ 'pour' (imperative, 2nd singular) is an example for a CQ-triplet. The distinction between long and overlong vowels and consonants is orthographically not marked; they are both represented by double letters, e.g. <aa> or <ll>. An exception are the plosives, where CQ1 is represented by /b d g/, CQ2 by /p t k/ and CQ3 by /pp tt kk/. Monosyllabic words are always overlong, e.g VQ3 *kool* [koːːl] 'school' (nominative singular), or they are part of a larger phonological unit and therefore fused with the neighbouring words (Lehiste, 1997a).

Segment durations alone are insufficient for the description of the Estonian quantity contrast. Instead, it was found that the differences between the three quantity degrees can be better described with the help of the so-called duration ratio between the rhymes of the first and the second syllable of the disyllabic foot of the form $C_1V_1C_2V_2$ (Lehiste, 1960). Typical ratios are 2:3 for Q1, 3:2 for Q2 and 2:1 for Q3 as found first by Lehiste (1960) and later by others (e.g. Eek, 1974; Krull, 1991, 1992; Liiv, 1961). The duration ratios were also found to be stable in spontanous speech (Asu et al., 2009; Krull, 1993; Lippus, 2010).

Besides these temporal differences, VQ-words differ in their tonal characteristics in both read and spontanous speech (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). To summarise the studies cited, it can be said that VQ3-words were characterised by an F0 peak early in the vowel of the first syllable, whereas the peak of VQ1- and VQ2-words was located towards the end of the vowel of the first syllable. In VQ3-words the fall was completed during the first syllable, and in VQ1- and VQ2-words the fall still took place in the second syllable. For CQ-words no comparable peak alignment differences were found (Krull, 1992; Lehiste, 1997b; Lippus et al., 2013). In the very recent study of Lippus et al. (2013) the turning point of a falling pitch accent was found in the vicinity of the offset of the vowel of the first syllable in both CQ2- and CQ3-words.

The quantity-dependent peak alignment differences described above only occur in falling H*+L accents (Asu et al., 2009), which is the most common of the three pitch accents of the Estonian pitch accent inventory in the

[2] Due to the foot isochrongy the vowel of the second syllable is half-long in Q1-words.

autosegmental-metrical analysis of Estonian intonation of Asu (2004, 2005)[3]. There are no peak alignment differences between the quantities neither in the also common low H+L* pitch accent (Asu and Nolan, 2007), nor in the rising L*+H pitch accent, which occurs frequently in colloquial speech (Asu and Nolan, 1999).

The perceptual distinction between the three quantity degrees in Estonian depends on an interaction between the temporal and tonal cues described above. On the segmental level, the quantity distinction is binary; listeners can only distinguish between short and long segments, but not between long and overlong segments (Eek and Meister, 1997, 2003). For the distinction between the three quantity degrees the whole disyllabic foot ist needed, expressed with the help of the duration ratios described above. Nevertheless, studies showed that the duration ratios alone are suitable for the distinction between the short and the long quantity degree, but often insufficient for the discrimination of the long and overlong quantity degree (e.g. Lehiste, 1970 1975, 1997a, 2003; Lehiste and Fox, 1992; Lippus et al., 2007, 2009, 2011; Salveste, 2011). For a successful discrimination of the long and overlong quantity degree additionally the tonal cues are used (e.g. Lehiste, 1970 1975; Lehiste and Fox, 1992; Lehiste, 1997a, 2003; Lippus et al., 2007, 2009, 2011; Salveste, 2011). Nevertheless, if the tonal cues are not present only the temporal cues are used for the discrimination between the long and overlong quantity degree, e.g. in whispered speech (Krull, 2001) or in the case of voiceless consonants (Lippus and Pajusalu, 2009; Lippus et al., 2009). The discrimination of the short and long quantity degree depends not only on temporal aspects but also on the intrinsic length of the vowels (Meister et al., 2011).

In general, Estonian has quite a free word order. Nevertheless, typical and frequent word orders are Subject-Verb-Object and Object-Verb-Subject (Lindström, 2006). Word order is used to express information structure, and the beginning and the end of a sentence can be focus positions (Lindström, 2006; Tael, 1988). A word in an unmarked focus position is usually placed at the end of a sentence, whereas a word in contrastive focus is located at the beginning of the sentence (Lindström, 2006). Nevertheless, two recent studies showed that information structure in Estonian is not only expressed by syntax, but also by prosody (Sahkai et al., 2013; Salveste, 2013).

[3]For a description of the AM model of Estonian intonation see chapter 1.2.2.

## 1.2 Intonation

This dissertation is concerned with research within an autosegmental-metrical (AM) framework of intonational phonology. Therefore firstly a general introduction into intonation models, containing a more detailed description of AM models, is given, and secondly the AM model of Estonian intonation is presented. The final part of this section is an overview of tonal alignment research since this dissertation is concerned with peak alignment.

### 1.2.1 Intonation Models

Two different approaches for the description of intonation have been developed over the years: contour-based models (e.g. Crystal, 1969; Halliday, 1967; Kingdon, 1958; 't Hart et al., 1990) and target models (e.g. Ladd, 1996; Pierrehumbert, 1980; Pike, 1945; Trager and Smith, 1951; Wells, 1945). The models of the so-called British school (e.g. Crystal, 1969; Halliday, 1967; Kingdon, 1958), the IPO approach for Dutch (e.g. 't Hart et al., 1990) or the model of Kohler (1991) for German are contour-based and describe intonation with the help of dynamic pitch movements, i.e. rises and falls or combinations of them. In target models pitch contours are described with a set of different level tones which are connected with certain points of the segmental string. In the earlier target models a set of four tones (Low, Mid, High and Overhigh) was used (Pike, 1945; Trager and Smith, 1951; Wells, 1945) whereas newer concepts such as the autosegmental-metrical (AM) theory of intonational phonology only uses two tones: High and Low (e.g. Ladd, 1996; Pierrehumbert, 1980). Since the current dissertation uses an AM model, in the following only the AM theory is described further (for a comparison of different contour models of the British school see e.g. Lecumberri (1997), and for a comparison of contour-based and target models see e.g. Grice and Baumann (2007) Ladd (2008) or Nolan (2006).

The AM theory of intonational phonology (e.g. Beckman and Pierrehumbert, 1986; Ladd, 1996; Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988), assumes that an intonational contour consists of different tonal targets which are connected to each other by transitions. The foundations for the AM approach were the dissertations of Bruce (1977), Liberman (1975) and particularly Pierrehumbert (1980). In later studies (e.g. Beckman and Pierrehumbert, 1986; Pierrehumbert and Beckman, 1988) the system was refined, and its main principles were condensed and extensively described in Ladd

(1996). The AM theory rests on four main principles (Ladd, 2008)[4]: Firstly, the tonal structure consists of different tonal events which are connected with each other by transitions. In intonation languages such as English two types of events can be distinguished: pitch accents, which are aligned with prominent syllables, and edge tones, which are associated with prosodic boundaries. The second principle of the AM theory implies that there is a distinction between pitch accents and stress. Pitch accents are intonational events that can be aligned with metrically strong syllables, but they do not serve as a direct acoustic cue for stress. According to the third principle, pitch accents and edge tones can be modelled in many languages with two levels (targets): a low target (L) and a high target (H). In languages such as English a monotonal pitch accent consists of either a low target or a high target, whereas a so-called bitonal pitch accent is formed by a combination of both. A bitonal pitch accent contains a starrred tone (H* or L*), which is usually associated with the stressed syllable and can be preceded (leading tone) or followed (trailing tone) by an L or H. The edge tones at phrase or sentence boundaries are either low (L) or high (H). The fourth principle of the AM theory concerns global trends. Global trends, such as declination, are attributed to localised and repeated changes in the scaling of low and high tones in the AM theory.

Within the AM framework, the ToBI (Tone and Break Index) transcription system is the most common system for transcribing intonation. It was first developed for the transcription of American English (Beckman and Ayers Elam, 1997; Silverman et al., 1992) and later adopted to several other languages (e.g. Arvaniti and Baltazani, 2005 for Greek; Grice et al., 2005 for German; Jun, 2000 for Korean; Prieto, in press for Catalan; Venditti, 2005 for Japanese).

### 1.2.2 The AM Model of Estonian Intonation

Estonian intonation was firstly described within an autosegmental-metrical (AM) framework by Asu (2004). This subsection summarises the main aspects of the Estonian AM model.

The pitch accent inventory of the Estonian AM model consists of two monotonal (H*, L*) and four bitonal (H*+L, L*+H, H+L*, H+!H*) pitch accents (Asu, 2004, 2005).

[4] Ladd (2008) is the second edition of Ladd (1996).

The AM model of Estonian uses only intonational phrases, but no intermediate phrases as the ToBI transcriptions for other languages (e.g. Beckman and Ayers Elam, 1997 for English). For the marking of the boundary tones of the intonational phrases the Estonian AM system provides two possibilities: Firstly, a boundary is left unspecified if there is no pitch movement at the boundary (%). This is the default case for Estonian (Asu, 2004, 2005). Secondly, in spontanous speech a rising movement near the boundary is possible which is marked with an H% (Asu, 2004).

The falling H*+L pitch accent is the most frequent pitch accent of Estonian in both read and spontanous speech, and it can occur in prenuclear and nuclear position (Asu, 2004, 2005). The H+L* pitch accent, a low tone preceded by a high tone that can appear in both nuclear and prenuclear position, is also common in read and spontanous speech (Asu, 2004, 2005; Asu and Nolan, 2007). H+L* pitch accents were only found in statements, never in questions (Asu and Nolan, 2007). The downstepped pitch accent H+!H* only occurs in spontanous speech, but not in read speech (Asu, 2004). Rising accents have been found quite frequently in spontaneous speech (Asu, 2004, 2006; Keevallik, 2003), even though Estonian rises are reported to be uncommon since questions in Estonian are usually realised with falling intonation (Asu, 2004). Three different rising pitch accents can be distinguished: L*+H %, L* H% and H* % which can occur in nuclear, but not in prenuclear position (Asu, 2004, 2005). Asu (2006) found a connection between rising intonation and discourse function in Estonian. Rising intonation often occured on short discourse markers such as *jaa* or *mhmh* (Asu, 2006). If the discourse marker was used as a turn-holder, a L*+H % pitch accent was used, and in the case of a turn-yielding function the L* H% pitch accent predominated (Asu, 2006).

A finite-state grammar (Figure 1.1) shows how the pitch accents and the boundary tones can be combined to different intonational tunes (Asu, 2004, 2005). A sequence of H*+L pitch accents followed by an unmarked boundary is the most widespread tune in read and spontanous speech (Asu, 2004, 2005). Other tunes consist of downstepped accents (H+!H*), low nuclei (H+L*), and rising accents, partially combined with the H*+L pitch accent (Figure 1.1). Furthermore the finite-state grammar shows that it is necessary in Estonian to distinguish between prenuclear and nuclear accents, because the rising pitch accents only occur in nuclear position (Asu, 2004, 2005, 2006).

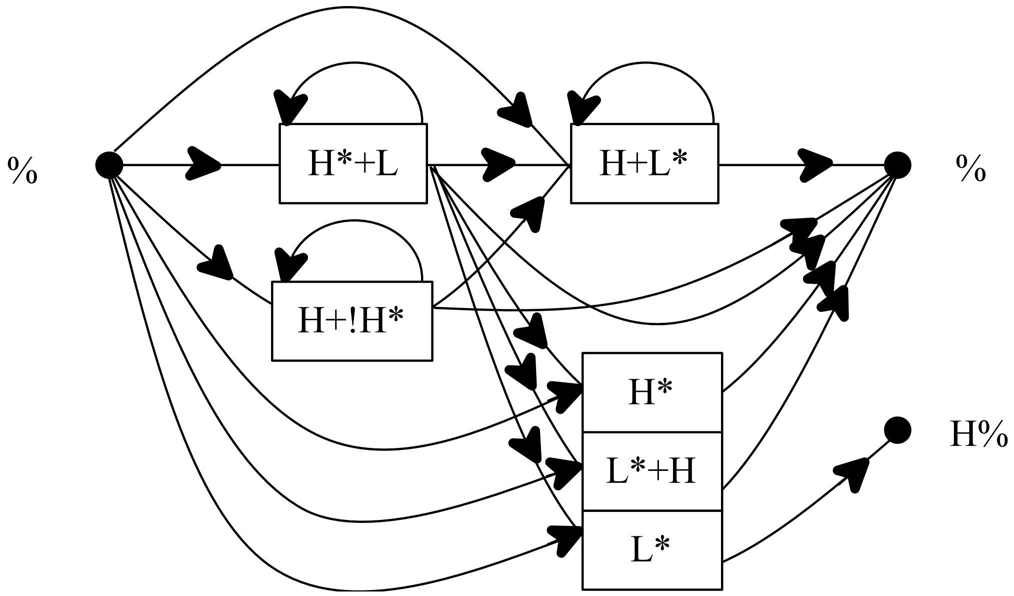


Figure 1.1: Finite-state grammar of Estonian intonation after Asu (2004, pg. 131)

### 1.2.3 Tonal Alignment

Tonal alignment is the temporal coordination of tonal targets with the segmental string (Ladd, 2008). Tonal alignment itself is theory-neutral, but nevertheless it fits very well with the basic assumption of the autosegmental-metrical (AM) theory of intonational phonology that pitch contours consist of strings of tonal targets which are associated with the segmental string in a clearly defined way (Ladd, 2008). Tonal alignment is regarded as a phonetic dimension; however, phonological distinctions can be derived from it (Ladd, 2008).The difference between the phonetic and the phonological dimension of tonal alignment is reflected in the terms *alignment* and *association*. While association is an abstract phonological property, alignment is a phonetic property; i.e. a tonal target can be associated e.g. with the stressed syllable of a word, but it is aligned early or late in this particular syllable or even a bit before or after it. The point of the segmental string a tonal target is aligned with can be the edge of a phonological domain, e.g. the beginning of the accented syllable, but it is not necessarily a phonological domain. An example for a non-phonological domain is e.g. '15 ms into the unstressed vowel' (Ladd, 1983, 2008).

A very relevant early work of tonal alignment is the dissertation of Bruce (1977) showing that the distinction between Accent 1 and Accent 2 in Swedish is a matter of the alignment of the peak with the stressed syllable. Since Bruce's work on Swedish, numerous studies investigating the influence of prosodic features on the tonal alignment of starred and unstarred tonal targets have been published. For an introduction and overview see e.g. Ladd (2008), D'Imperio (2006) and the studies cited in this subsection.

It was shown that several prosodic features can influence the alignment of tonal targets, such as time pressure due to the righthand prosodic context, i.e. the proximity to prosodic boundaries or other pitch accents due to a variation of the number of unstressed syllables following the pitch accented word (e.g. Arvaniti and Ladd, 1995; Arvaniti et al., 1998 for Greek; De la Mota, 2005; Prieto, 2005; Prieto et al., 1995 for Spanish; Gili Fivela, 2002 for Italian; Mücke and Hermes, 2007; Rathcke and Harrington, 2007 for German; Silverman and Pierrehumbert, 1990 for English), syllable structure (e.g. D'Imperio, 2000; D'Imperio et al., 2007; Gili Fivela and Savino, 2003 for Italian; Hellmuth, 2006 for Arabic; Ishihara, 2003 for Japanese; Prieto and Torreira, 2007; Prieto, 2009; Prieto et al., 1995 for Spanish; Welby and Lœvenbruck, 2005, 2006 for French), the position of the word accent (e.g. Arvaniti et al., 1998 for Greek; Prieto et al., 1995 for Spanish) or phonological vowel length (e.g. Asu et al., 2009; Lippus et al., 2013; Plüschke, 2011 for Estonian; Ladd et al., 2000; Schepman et al., 2006 for Dutch).

Nevertheless, there are also studies showing that tonal targets were consistently aligned with specific points of the segmental string regardless of prosodic factors such as the segmental composition of the words or speaking rate (e.g. Arvaniti et al., 1998; Igarashi, 2004; Ladd et al., 1999; Xu, 1998). This phenomenon is called segmental anchoring, a term which was first introduced by Ladd et al. (1999), and inspired by the findings of Arvaniti et al. (1998) that in Greek both the beginning and the end of a rising pitch accent were aligned with specific landmarks of the segmental string regardless of the durational composition of the segments. Since then the search for segmental anchors has been object of a great number of studies in various languages (e.g. Atterer and Ladd, 2004; Kleber and Rathcke, 2008; Mücke et al., 2009 for German; Dalton and Ní Chasaide, 2005 for Irish; D'Imperio et al., 2007 for Italian and French; Igarashi, 2004 for Russian; Dilley et al., 2005; Ladd et al., 2000; Schepman et al., 2006 for Dutch; Ladd et al., 1999, 2009 for

English; Prieto and Torreira, 2007; Prieto, 2009 for Spanish, Welby, 2006 for French; Xu, 1998 for Mandarin Chinese).

The studies presented so far concerned speech production, but it was shown that tonal alignment does not only concern speech production, but also speech perception; it was found that alignment differences signaled semantic and pragmatic differences (e.g. Dilley and Brown, 2007; D'Imperio, 2000; D'Imperio et al., 2007; Kohler, 1987, 2005; Niebuhr, 2007; Pierrehumbert and Steele, 1989).

# Chapter 2

# Peak Alignment and post-focal unstressed syllables

## 2.1 Introduction

The aim of this chapter was to study the influence of the upcoming sentence boundary on the peak alignment of Estonian falling pitch accents and its interaction with quantity-dependent peak alignment differences. This was done via an experiment varying the number of intervening unstressed syllables between the target word and the sentence boundary. Both vowel (VQ) and consonant (CQ) words have been investigated.

Several studies have shown for various languages that the alignment of the high target of nuclear and prenuclear pitch accents is sensitive to the extent of the post-focal material and more specifically that the high target of accented words is aligned earlier in the word, the fewer the number of post-focal syllables that follow it. This peak shift to the left was found e.g. for English (Silverman and Pierrehumbert, 1990; Steele, 1986), for different varieties of German (Möbius and Jilka, 2007; Mücke and Hermes, 2007; Peters, 1999; Rathcke and Harrington, 2007), for Greek (Arvaniti and Ladd, 1995; Arvaniti et al., 1998), for Castilian Spanish (De la Mota, 2005) and for Mexican Spanish (Prieto et al., 1995). However, the low target of the pitch accents was found to be rather invariant, its alignment was not affected by the number of post-focal unstressed syllables (Arvaniti and Ladd, 1995; Arvaniti et al., 1998; Prieto et al., 1995). Nevertheless, the results of Rathcke and Harrington (2007) are different. They found a peak shift not only for the high target but also for the low target of a bitonal pitch accent.

Steele (1986) (described in more detail in Silverman and Pierrehumbert (1990)) investigated nuclear H* pitch accents in English read speech. The number of postnuclear syllables was either zero or three. They reported an earlier peak for the test words which were followed by none postnuclear syllables. The peak location was expressed as a function from the vowel onset to the peak in proportion to the vowel duration.

Silverman and Pierrehumbert (1990) reported a peak shift to the left for English prenuclear H* accents of read speech of two speakers, if the number of postfocal syllables decreased. The number of postfocal syllables extended from none to three. They found the peak shift for both absolute and proportional measures, claiming that the proportional measures had more regular patterns. For the absolute measures the peak was measured as the duration between the vowel onset and the peak. For the proportional measures this duration was divided by the rhyme length.

Peters (1999) analysed German H*, L+H* and L*+H nuclear accents of a corpus of spontaneous speech for two dialects of German: Berlin and Hamburg German. The number of postnuclear syllables was calculated, but the variation of the number was not reported. The peak location was calculated in absolute measures as the duration from the beginning of the syllable to the high target. In both varieties there was a peak shift to the left of the high target of the pitch accents, if the number of postnuclear syllables decreased.

Möbius and Jilka (2007) analysed German H*L nuclear accents in a read-speech corpus. The pitch accents were categorised as having the pitch accent either on the last syllable of the word or earlier. An earlier peak placement was found for the words where the pitch accent was later in the word, in other words if there were fewer unstressed syllables following the accented syllable.

Mücke and Hermes (2007) analysed rising nuclear accents of Viennese German of two speakers. The test words were in phrase-medial and phrase-final position in contrastive focus. Therefore in phrase-final words the accented syllable was followed by one unstressed syllable and in phrase-medial words by four unstressed syllables. The high target of the pitch accent was earlier in phrase-final in comparison to phrase-medial words, in other words it was earlier if there were fewer unstressed syllables following the accented syllable. The peak was expressed in absolute measures and all possible vowel and consonant onsets and offsets were used as landmarks for measuring the peak location.

Rathcke and Harrington (2007) found an earlier peak alignment for the high targets of German H* and H*L nuclear pitch accents if the number of postnuclear syllables decreased. The number of postnuclear syllables was either one, none or two. The low target of the H*+L pitch accent was also aligned earlier if there were fewer postnuclear unstressed syllables. The landmark for determining the peak was the onset of the accented syllable, and the peak alignment was expressed in proportion to the syllable duration.

Arvaniti and Ladd (1995) investigated Greek prenuclear rising accents of three speakers. While the initial low F0 target was stable, the high tone target was aligned earlier if the number of postnuclear syllables decreased. The number of unstressed syllables following the prenuclear accent varied between one and five. The peak was expressed in absolute measures as the duration between the onset of the accented syllable and the peak.

The results of the third experiment in Arvaniti et al. (1998) confirmed the results of Arvaniti and Ladd (1995) with an analysis of Greek prenuclear rising accents of three speakers. The number of unstressed syllables following the accented syllable was either one or two. The speakers formed two groups: in group one, consisting of two speakers, the alignment of the high target was not affected by the number of unstressed syllables, while in the other group there was a peak shift to the left, if there were fewer unstressed syllables after the accented syllable. The peak was expressed with absolute measures as the duration from the onset of the first vowel to the high target and as the duration between the low and the high target.

De la Mota (2005) analysed Castilian Spanish rising prenuclear accents of two speakers and varied the number of unstressed syllables between the prenuclear and the following pitch accent. In the first set of test sentences the number of unstressed syllables was either one or two and in the second set either two or three. If the number of syllables between the two accents decreased, the high target of the rising pitch accent was timed earlier.

Prieto et al. (1995) investigated speech material of two Mexican Spanish speakers. They analysed H* accents and varied the number of syllables between the accented and the following stressed syllable, with the number of unstressed syllables varying between zero and four. For both speakers they found a tendency for an earlier peak alignment if the number of unstressed syllables decreased. The start of the rise for the H* accent was found to be anchored consistently around the syllable onset and was not influenced by the number of unstressed syllables following the accented syllable.

Despite the studies described above, there is one study for Irish in which the number of unstressed syllables following a nuclear pitch accent had no effect on the peak alignment (Dalton and Ní Chasaide, 2005). Nevertheless, the authors do not exclude the possibility that indeed there was an effect, because the onset consonant, where the peak was probably located, was voiceless.

In all studies cited above regarding the influence of unstressed syllables on the peak alignment the peak of both nuclear and prenuclear accents was timed earlier if the number of post-focal unstressed syllables decreased (Arvaniti and Ladd, 1995; Arvaniti et al., 1998; De la Mota, 2005; Möbius and Jilka, 2007; Mücke and Hermes, 2007; Peters, 1999; Prieto et al., 1995; Silverman and Pierrehumbert, 1990; Steele, 1986). For Irish such an effect was not found (Dalton and Ní Chasaide, 2005).

In the current study, the effect of the number of postnuclear unstressed syllables was tested for Estonian. Both vowel (VQ) and consonant (CQ) quantity words produced with a falling nuclear H*+L pitch accent were analysed in two contexts: in the short tail context there were no unstressed syllables following the target word and in the long tail context there were two unstressed syllables after the target word. It was hypothesised that the peak is timed earlier in short tail words.

Peak alignment in Estonian always must be seen in connection with quantity since vowel quantity is not only signaled temporally (Eek, 1974; Krull, 1991, 1992; Lehiste, 1960; Liiv, 1961), but also tonally (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). The second aim of this chapter is therefore to study the interaction between the quantity-dependent peak alignment differences with the peak alignment differences due to the number of unstressed syllables.

Disyllabic VQ-words ($C_1V_1C_2V_2$) differ in the realisation of the pitch contour in both read and spontanous speech(Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). VQ3-words were characterised by an F0 peak early in the vowel of the first syllable, whereas the peak of VQ1- and VQ2-words was located towards the end of the vowel of the first syllable. In VQ3-words the fall was completed during the first syllable and in VQ1- and VQ2-words the fall still took place in the second syllable. For CQ-words no comparable peak alignment differences between the quantity degrees were reported (Krull, 1992; Lehiste, 1997b;

Lippus et al., 2013). The only difference found between CQ2- and CQ3-words was a larger F0 fall during the intersyllabic consonant in CQ3-words than in CQ2-words (Krull, 1992). In the very recent study of Lippus et al. (2013) the turning point of the high target of an H*+L pitch accent of CQ-words was located in the vicinity of the $V_1$ offset: in CQ2-words the turning point was found slightly after and in CQ3-words a bit before the $V_1$ offset. VQ1- and CQ1-words were analysed together and their turning points were located slightly after the $V_1$ offset. Lippus et al. (2013) found similarities in the absolute alignment of the turning point in relation to the $V_1$ onset: in CQ2-, CQ3-, VQ2- and VQ3-words the turning points were realised after a similar duration from the $V_1$ onset.

Since the quantity-dependent peak alignment differences of VQ-words only occured in H*+L pitch accents (Asu et al., 2009), the current study concentrated on nuclear H*+L accents to investigate the interaction between quantity and post-focal syllable material dependent peak alignment differences. It was proposed that the quantity specific peak alignment, namely an early peak for VQ3-words and a late peak for VQ1- and VQ2-words is maintained regardless of the peak shift due to the varied number of post-focal syllables. Additionally, the influence of post-focal syllables on CQ-words was investigated, and it was proposed that similarily as in Krull (1992), Lehiste (1997b) and Lippus et al. (2013) there are no peak alignment differences due to quantity in CQ-words.

The following hypotheses were tested:

H1) The high target of an H*+L accent is timed earlier in short tail words in all three quantity degrees in both vowel and consonant quantity.

H2) Regardless of H1 the quantity-dependent typical peak alignment of VQ-words is still maintained.

H3) There are no peak alignment differences due to quantity in CQ-words.

## 2.2 Method

### 2.2.1 Subjects and materials

Nine speakers of Standard Estonian participated in the experiment including six female and three male subjects. They were aged between 21 and 31 years. One female subject was a phonetician, but she was not familiar with the

Table 2.1: Target words with vowel quantity.

| **VQ1 (short)** (words in gen. and nom. sg.) | **VQ2 (long)** (words in gen. sg.) | **VQ3 (overlong)** (words in part. sg.) |
|---|---|---|
| /lɤmɑ/ ('made-up word') | /lɤːmɑ/ ('flame') | /lɤːːmɑ/ ('flame') |
| /mimi/ ('Mimi') | /miːmi/ ('mime') | miːːmi/ ('mime') |
| /mini/ ('very small, miniskirt') | /miːni/ ('mine') | /miːːni/ ('mine') |
| /mɤna/ ('spell') | /mɤːna/ ('low tide') | /mɤːːna/ ('low tide') |

purpose of the experiment. The recordings were made with a high quality head set microphone in a professional recording studio in Tartu, Estonia. The subjects were paid for their participation. The material consisted of four target word triplets for vowel quantity (Table 2.1) and four triplets for consonant quantity (Table 2.2). The VQ-triplets only differed in the duration of the vowel of the first syllable. The CQ-triplets differed in the duration of the intersyllabic consonant. With the exception of one item, all were existing lexical items in Estonian.

Carrier phrases were designed to elicit a falling H*+L pitch accent on the target word (Table 2.3). The target words were embedded in narrow focus position in two different carrier phrases: the first carrier phrase contained no unstressed syllables after the target word and in the second one two unstressed syllables followed the target word (henceforth, these contexts will be referred to as short and long tail contexts respectively). The carrier phrases for the long and overlong quantity had to be constructed differently because of differences in the target words' grammatical category. Due to grammatical differences also the carrier sentence for the target word *nimme* for the overlong quantity degree had to be constructed differently.[1]

The carrier sentences were presented as answers to questions to elicit narrow focus on the target word. For the recordings, the recording pro-

[1] *Sa lõikad* *nimme*. ('You - cut - into the loin'.) - *Sa lõikad* *nimmegi ju*. ('You - cut - into the loin, too - indeed.')

Table 2.2: Target words with consonant quantity.

| **CQ1 (short)** (words in gen. and nom. sg.) | **CQ2 (long)** (words in gen. sg.) | **CQ3 (overlong)** (words in part. sg.) |
|---|---|---|
| /hɑlɑ/ ('moaning') | /hɑllɑ/ ('frost') | /hɑlːlɑ/ ('frost') |
| /hæli/ ('Häli') | /hælli/ ('cradle') | /hælːli/ ('cradle') |
| /linɑ/ ('linen') | /linnɑ/ ('city') | /linːnɑ/ ('city') |
| /nime/ ('name') | /nimme/ ('loin', only nom. sg.) | /nimːme/ ('loin', only ill. sg.) |

Table 2.3: Carrier sentences for eliciting H*+L pitch accents on the target word.

| **Quantity degree** | **Short tail** | **Long tail** |
|---|---|---|
| Short/Long | *Sa leidsid* ____ *.* ('You - found - ____ .') | *Sa leidsid* ____*gi ju.* ('You - found - ____ , too.') |
| Overlong | *Sa nägid* ____ *.* ('You - saw - ____ .') | *Sa nägid* ____*gi ju.* ('You - saw - ____ , too.') |

gram *SpeechRecorder* (Draxler and Jänsch, 2004) was used, which prompted first the question and then the corresponding answer (which was the carrier sentence) on the screen of a notebook computer. Only the answers to the questions were recorded. Each carrier sentence was read once. Not every target word was read by every subject. The carriers sentences were read triplet by triplet. Within each triplet, first the short and the long tail carrier sentence containing the Q1-word was read, and then the carrier sentences with the Q2- and the Q3-words. After the recording of the VQ-words, the subjects were asked to read carrier sentences for eliciting H+L* and L*+H pitch accents containing the same VQ-words. Finally, the whole procedure was repeated for the CQ-words. The carrier sentences for the H+L* and the L*+H pitch accent are not presented here since they were not used for the analysis of the current study.

After eliminating recordings containing errors (e.g. the subject produced the vowel with a different quantity than the one indicated), for the H*+L accent a total of 176 utterances with vowel quantity and 174 utterances with consonant quantity remained for the analysis.

### 2.2.2 Analysis

The data were automatically divided into phonetic segments using the Munich Automatic Segmentation System MAUS (Schiel, 1999) and subsequently manually corrected with the help of the labeling tool Praat (Boersma and Weenink, 2010). The F0 contours were determined by means of the Schaefer-Vincent periodicity detector (Schaefer-Vincent, 1983) in Emu (Harrington, 2010). The utterances were checked in order to determine whether the intended pitch accent had actually been produced or not. An auditory and visual analysis of the utterances' pitch accents showed that in vowel quantity 81.25% of the target words were produced with an H*+L accent and 18.75% with an H+L* accent. In consonant quantity 74.71% of the target words were realised with an H*+L accent and 25.29% with an H+L* accent.

For the analysis of the H*+L pitch accents only those H*+L accents were used, which consisted of a clear rising-falling sequence. This was necessary to allow the algorithm described below to detect the F0 maximum automatically. Furthermore some of the CQ-words had to be excluded, where the peak extended in a voiceless initial consonant. A total of 103 utterances of vowel quantity and 99 utterances of consonant quantity remained for the analysis.

For the automatic detection of the peak the F0 contours of the target

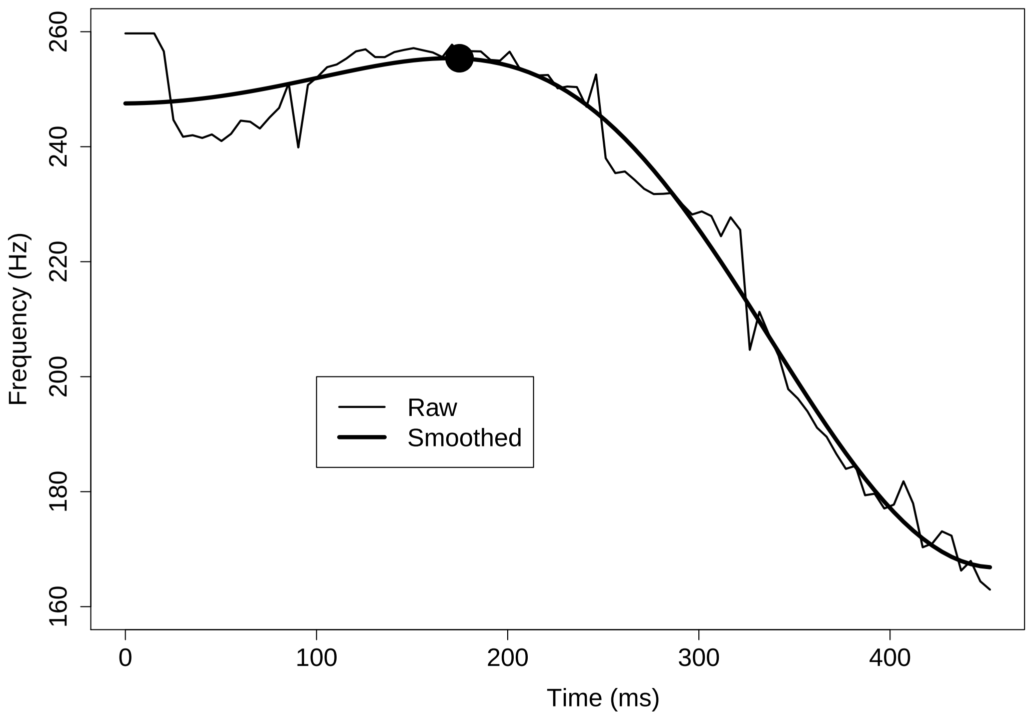


Figure 2.1: The raw and the smoothed contour of the long quantity target word *mimi* spoken by the speaker KK.

words were smoothed in R (Figure 2.1) using the discrete-cosine-transformation (DCT) following a similar procedure used in Rathcke and Harrington (2006). Before smoothing, errors in pitch-tracking were deleted and approximated to the neighbouring F0 values. The approximated contour was smoothed with the help of the DCT using the first four coefficients. A DCT decomposes a signal into cosine waves. The original signal can be obtained by summing up all the obtained cosine waves. If only the first few of the cosine waves are summed up, the result is a smoothed signal (Harrington, 2010). A pilot study showed that the smoothing with only three coefficients was insufficient, because it smoothed the contour too much. The smoothing with four and five coefficients was similar. Therefore four coefficients were found to be optimal for the purposes of peak detection.

After the smoothing, the F0 maximum of the smoothed contour of an H*+L accent was detected automatically and checked manually (Figure 2.1).

The alignment of the peak was expressed in absolute and proportional measures. The absolute peak alignment, $T_{abs}$ was calculated from

(1) $T_{abs} = T - V_{on}$

where T is the time of the tone target (the time of occurrence of the starred tone H* in the bitonal H*+L accent) and where $V_{on}$ is the acoustic vowel onset of $V_1$.

The proportional peak alignment $T_{VCprop}$ was calculated relative to the acoustic onset of $V_1$ and offset of $C_2$ from (2)

(2) $T_{VCprop} = T_{abs}/(C_{off} - V_{on})$

where $V_{on}$ is the acoustic onset of $V_1$ and $C_{off}$ is the acoustic offset of $C_2$.

Furthermore the proportional peak alignment $T_{Vprop}$, was calculated relative to the acoustic onset and offset of $V_1$ from (3)

(3) $T_{Vprop} = T_{abs}/(V_{off} - V_{on})$

where $V_{on}$ is the acoustic onset of $V_1$ and $V_{off}$ is the acoustic offset of $V_1$ (see also Asu et al., 2009; Silverman and Pierrehumbert, 1990; Steele, 1986 for a similar procedure).

The R-software combined with the EMU/R-package (Harrington, 2010) were used for all subsequent analyses. The statistical analyses reported below were carried out mainly using mixed models in R. Significance values for the different factors were achieved by comparing a mixed model with and without the independent factor.

## 2.3 Results

Three hypotheses were tested regarding the peak alignment of H*+L accents. Firstly, the peak was hypothesised to be timed earlier in words with a short tail (H1). The second hypothesis (H2) was that despite this peak shift the quantity-dependent typical peak alignment should be maintained: VQ3-words should have an early peak at the beginning of the vowel of the

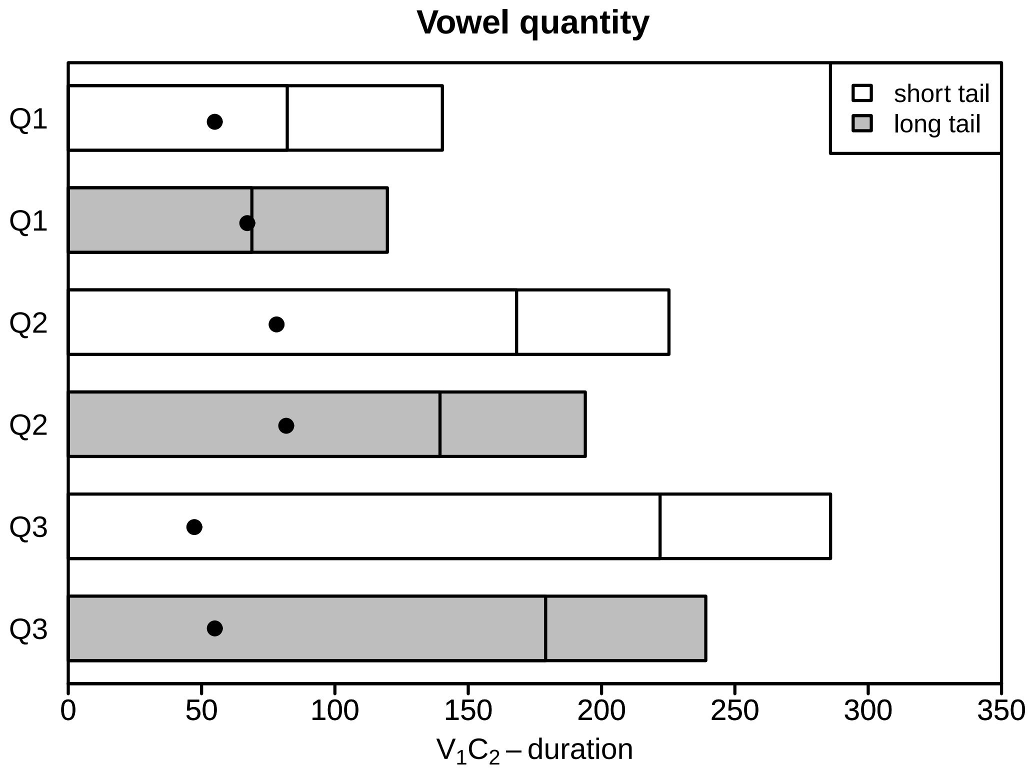


Figure 2.2: The bars show the $V_1C_2$-duration separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the short (white bars) and long tail (grey bars) context for VQ-words. The lines are the border between $V_1$ and $C_2$. The dots mark the mean value of $T_{abs}$.

accented syllable and VQ1- and VQ2-words a late peak at the end of the vowel. The third hypothesis (H3) proposed that there are no peak alignment differences due to quantity in CQ-words.

As Figure 2.2 shows, the mean value of $T_{abs}$ was timed earlier in the short tail context of VQ-words as proposed in hypothesis H1. The alignment of the median value of $T_{abs}$ for VQ-words was found to depend on the quantity degree (Figure 2.8). In VQ1- and VQ3-words the peak was earlier in the short tail context, but in VQ2-words the variation was higher and the peak was later than in the long tail context. For CQ-words both the mean and the median values of $T_{abs}$ were shifted leftwards in the short tail context (Figures 2.3 and 2.9). The proportional peak alignment $T_{VCprop}$ was earlier

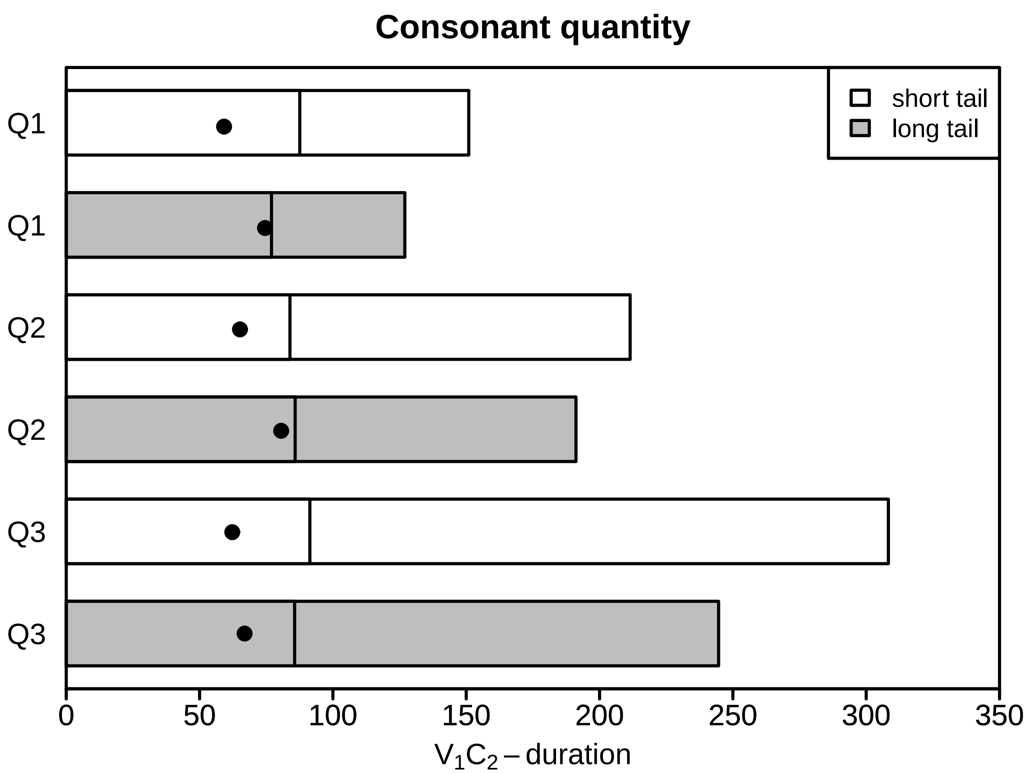


Figure 2.3: The bars show the $V_1$-duration separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the short (white bars) and long tail (grey bars) context for CQ-words. The lines are the border between $V_1$ and $C_2$. The dots mark the mean value of $T_{abs}$.

in the short tail context of both VQ- and CQ-words (Figures 2.4 and 2.5). Regarding the peak alignment of $T_{\mathrm{Vprop}}$ it was timed earlier in the short tail context of VQ1-, VQ2-, VQ3-, CQ1- and CQ2- words, but not in CQ3 words (Figures 2.6 and 2.7). In CQ3-words the peak alignment was similar in the short and the long tail context.

The purpose of hypothesis H2 was to test whether VQ1- and VQ2-words have a late and VQ3-words an early peak in the vowel of the first syllable despite the peak shift due to hypothesis H1. Figure 2.2 shows that VQ3-words indeed had an early peak and VQ1-words a late peak. However the peak of VQ2-words was not aligned at the end of the vowel, but rather in the middle of the vowel. The proportional peak alignment had the order VQ3 $<$ VQ2 $<$ VQ1, where $<$ denotes that the peak of VQ3 was timed to occur proportionally earlier than the peak of VQ2 and so on (Figures 2.4 and 2.6).

Hypothesis H3 proposed that there are no peak alignment differences due to quantity in CQ-words. As can be seen in Figure 2.3 the unit of observation is important if there are differences in the peak alignment of CQ1-, CQ2- and CQ3-words or not. The peak of CQ1-, CQ2- and CQ3-words was aligned with the end of the vowel of the first syllable (Figure 2.3). Therefore there were hardly any differences in the proportional peak alignment $T_{\mathrm{Vprop}}$ of CQ1-, CQ2- and CQ3-words (Figure 2.7). Differences in the proportional peak alignment occured between CQ1-, CQ2- and CQ3-words if not only the vowel ($V_1$) , but also the following consonant ($C_2$) was taken into acount. The proportional peak alignment $T_{\mathrm{VCprop}}$ of CQ-words had the order CQ3 $<$ CQ2 $<$ CQ1, similar to the alignment $T_{\mathrm{VCprop}}$ of VQ-words (Figure 2.6). Both the order of the peak alignment and the location of the peak within the $V_1C_2$-sequence were found to be similar between VQ- and CQ-words. In Q1-words the peak was located around the temporal midpoint of the $V_1C_2$-sequence, in Q2-words a bit before this midpoint and in Q3-words in the first third (Figures 2.2 and 2.3).

Mixed Models confirmed the observations for the hypotheses H1-H3 regarding the alignment of $T_{\mathrm{VCprop}}$ and $T_{\mathrm{abs}}$. The mixed models were carried out separately for VQ- and CQ-words. Dependent variables were $T_{\mathrm{VCprop}}$ and $T_{\mathrm{abs}}$ respectively. Quantity (three levels: short (Q1), long (Q2), overlong (Q3)), $V_1C_2$-duration and Tail (two levels: short, long) were factors. Speaker was a random factor.

The mixed model for VQ-words with $T_{\mathrm{VCprop}}$ as the dependent variable showed a significant effect for the Tail ($\chi^2[1] = 9.44$, $p < 0.001$) and Quantity ($\chi^2[2] = 16.28$, $p < 0.001$). The location of $T_{\mathrm{VCprop}}$ in VQ1- vs. VQ2-words

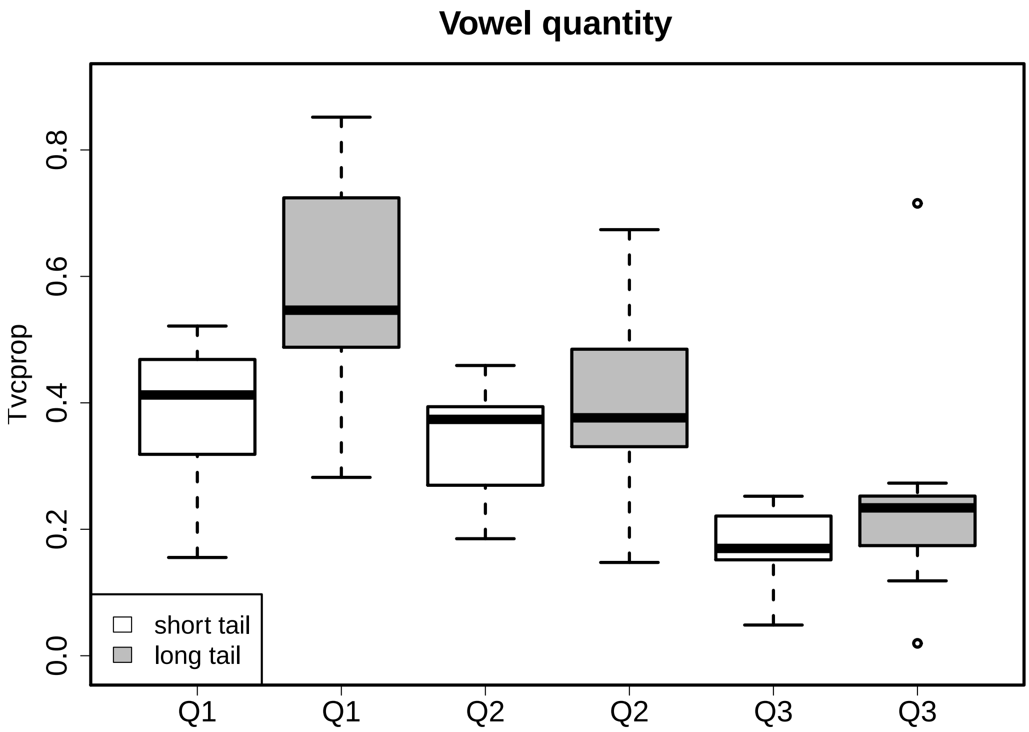


Figure 2.4: The proportional time for the tone target, $T_{VCprop}$, shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the short (white boxes) and long tail (grey boxes) contexts for VQ-words. In this figure (and all other figures containing boxplots in this dissertation) the midpoint of the box is the median, the top of the box the third quartile and the bottom of the box the first quartile. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box.

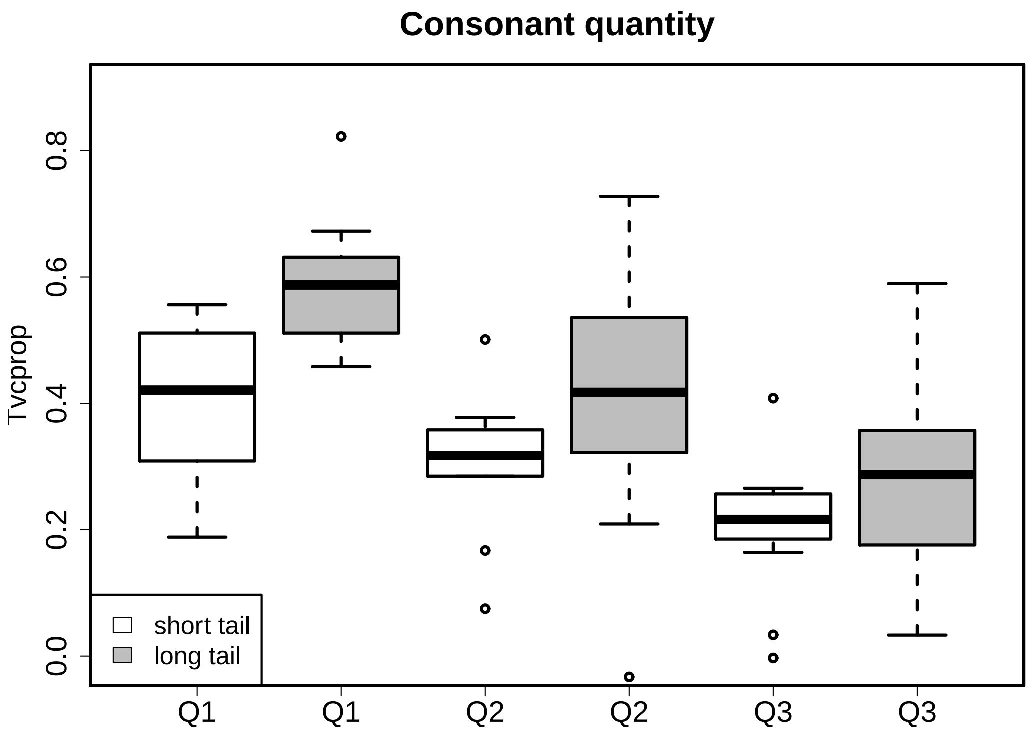


Figure 2.5: The proportional time for the tone target, $T_{VCprop}$, shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the short (white boxes) and long tail (grey boxes) contexts for CQ-words.

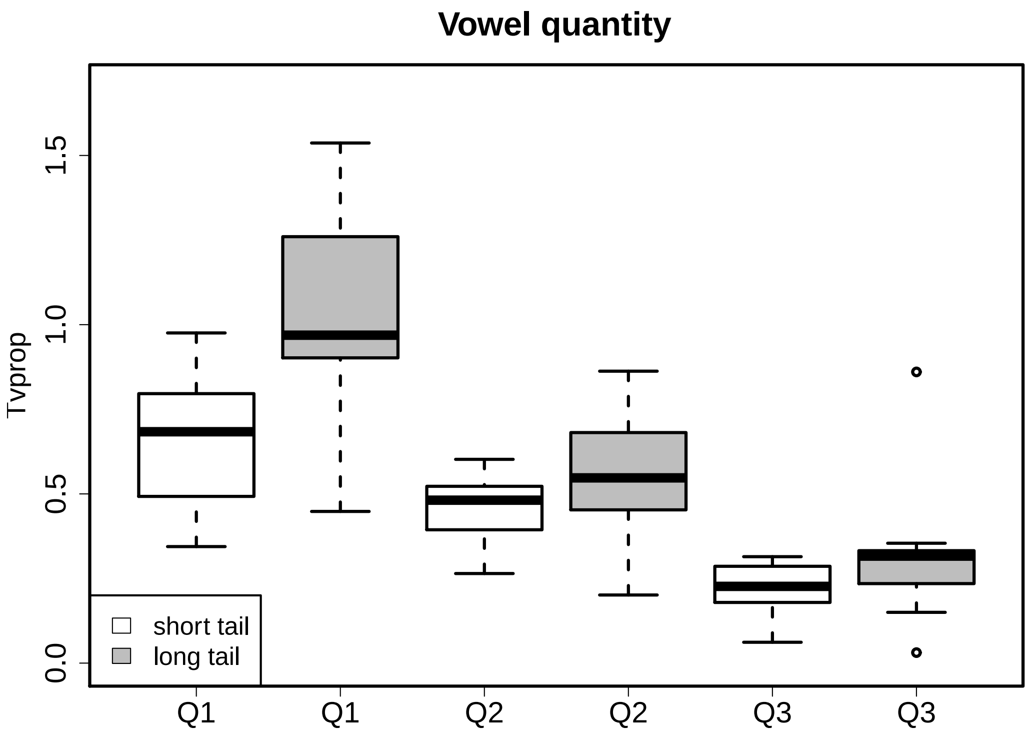


Figure 2.6: The proportional time for the tone target, $T_{\mathrm{Vprop}}$, shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the short (white boxes) and long tail (grey boxes) contexts for VQ-words.

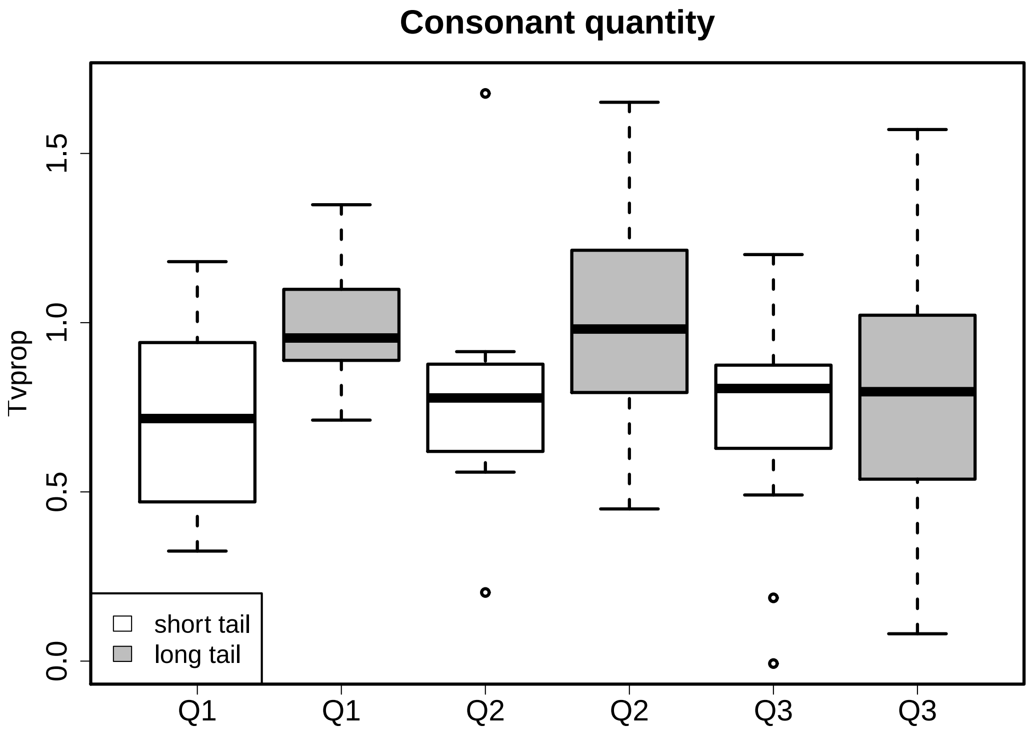


Figure 2.7: The proportional time for the tone target, $T_{\text{Vprop}}$, shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the short (white boxes) and long tail (grey boxes) contexts for CQ-words.

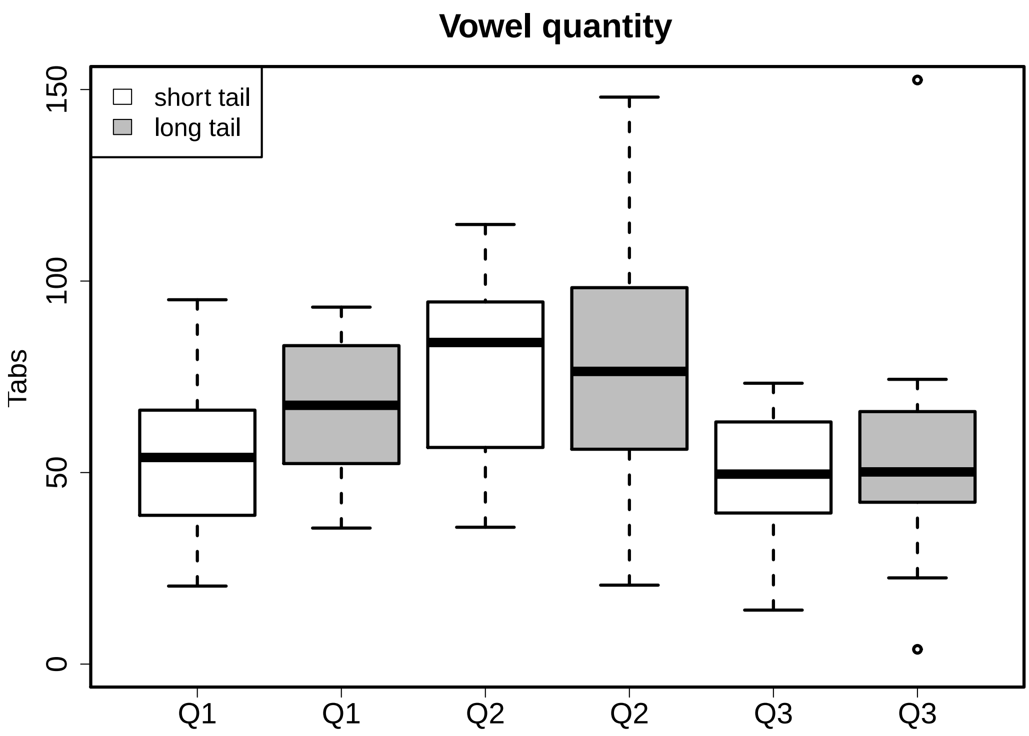


Figure 2.8: The absolute time for the tone target, $T_{abs}$, shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the short (white boxes) and long tail (grey boxes) contexts for VQ-words.

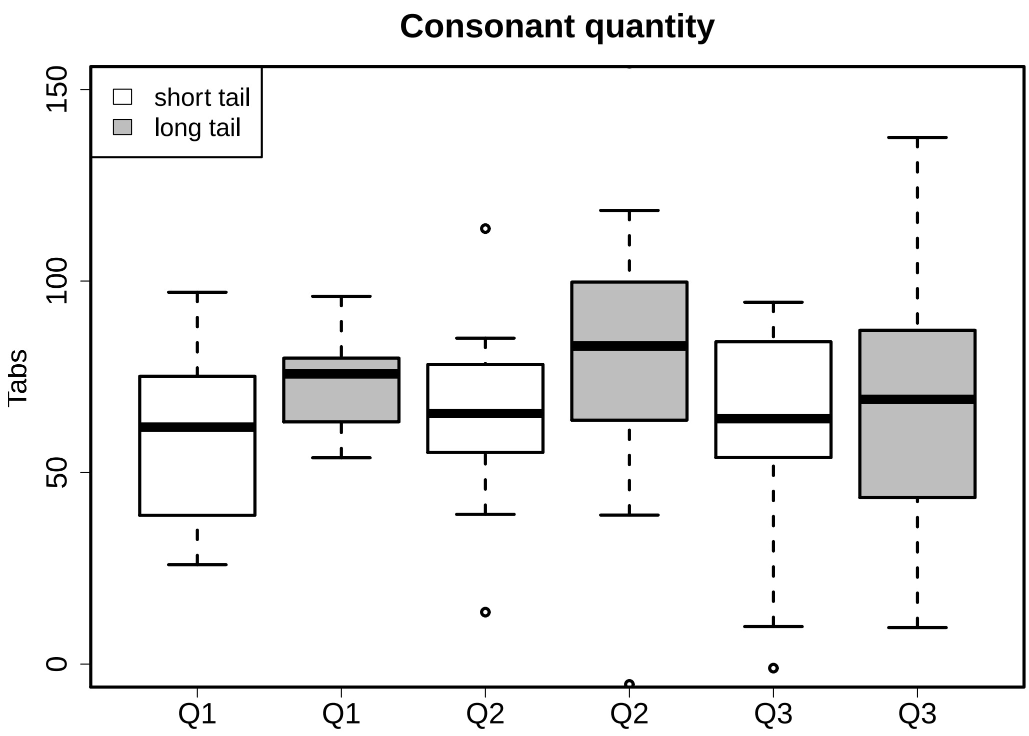


Figure 2.9: The absolute time for the tone target, $T_{abs}$, shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the short (white boxes) and long tail (grey boxes) contexts for CQ-words.

did not differ from each other, but so did it in VQ2- vs. VQ3-words ($z = 0.04$, $p < 0.001$) and in VQ1- vs. VQ3-words ($z = 0.07$, $p < 0.001$). The $V_1C_2$-duration did not have a significant influence on the alignment of $T_{VCprop}$. There was a significant triple interaction $V_1C_2$-duration x Quantity x Tail ($\chi^2[2] = 7.57$, $p < 0.05$).

In CQ-words the factors Tail ($\chi^2[1] = 11.76$, $p < 0.001$) and Quantity ($\chi^2[2] = 12.97$, $p < 0.01$) had a significant influence on the alignment of $T_{VCprop}$. The location of $T_{VCprop}$ was significantly different from each other in CQ1- vs. CQ2-words ($z = 0.04$, $p < 0.05$), in CQ2- vs. CQ3-words ($z = 0.04$, $p < 0.05$) and CQ1- vs. CQ3-words ($z = 0.06$, $p < 0.001$). The $V_1C_2$-duration did not have a significant influence on the alignment of $T_{VCprop}$.

The alignment of $T_{abs}$ in VQ-words was significantly influenced by Tail ($\chi^2[1] = 8.12$, $p < 0.01$), the $V_1C_2$-duration ($\chi^2[1] = 12.99$, $p < 0.001$) and Quantity ($\chi^2[2] = 38.22$, $p < 0.001$). The peak alignment $T_{abs}$ was not significantly different in VQ1- vs. VQ2-words. Nevertheless, the location of $T_{abs}$ was significantly different in VQ2- vs. VQ3-words ($z = 5.83$, $p < 0.001$) and VQ1- vs. VQ3-words ($z = 10.77$, $p < 0.001$).

The mixed model for the location of $T_{abs}$ in CQ-words showed a significant effect for Tail ($\chi^2[1] = 9.10$, $p < 0.01$), the $V_1C_2$-duration ($\chi^2[1] = 6.56$, p $< 0.05$) and Quantity ($\chi^2[2] = 8.99$, $p < 0.05$). There were no significant differences in the alignment of $T_{abs}$ in CQ1- vs. CQ2-words, but in CQ2- vs. CQ3-words ($z = 7.64$, $p < 0.01$) and CQ1- vs. CQ3-words ($z = 12.03$, $p < 0.05$).

## 2.4 Discussion

One of the main findings of this study was that the peak of a nuclear H*+L accent was timed earlier in short tail words. Proportional measures were found to be more stable than absolute measures. It was found that the peak alignment differences between VQ1-, VQ2- and VQ3-words were stable in the short vs. the long tail context. Furthermore it was shown that quantity differences, whether these are due to the vowel (VQ1 vs. VQ2 vs. VQ3) or the consonant (CQ1 vs. CQ2 vs. CQ3) were cued by proportional peak alignment within the $V_1C_2$-sequence and by its alignment with the offset of $V_1$ in VQ1- and CQ1-words, with a point slightly before the middle of the $V_1C_2$-sequence in VQ2- and CQ2-words and within the first third of the $V_1C_2$-sequence in VQ3- and CQ3-words.

The effect of the tail on the peak alignment is similar to that found for nuclear and prenuclear accents of non-quantity languages such as English (Silverman and Pierrehumbert, 1990; Steele, 1986), German (Möbius and Jilka, 2007; Mücke and Hermes, 2007; Peters, 1999; Rathcke and Harrington, 2007), Greek (Arvaniti and Ladd, 1995; Arvaniti et al., 1998) and Spanish (De la Mota, 2005; Prieto et al., 1995). It seems likely that time pressure is the reason for the peak shift in short tail words. There is less time to realise the pitch accent in words with a short tail, because there is less segmental material. To be able to still realise the pitch accent, the peak is shifted leftwards in short tail words. This leftwards-shift was achieved by two means: firstly the peak of short tail words was timed proportionally earlier, because of phrase-final lengthening of the word. Secondly there was a leftwards-shift also in absolute terms, despite in VQ2-words. The reason for VQ2-words being an exception could be the high variation in the data of VQ2-words.

Peak alignment can be expressed in both absolute and proportional terms, and the question arises which measurement is more stable for expressing the peak. Silverman and Pierrehumbert (1990) claimed that the proportional measures of the peak delay were more regular than the absolute measures. Nevertheless, Prieto et al. (1995) were able to obtain more consistent results for the absolute than for the proportional measurements. As in Silverman and Pierrehumbert (1990) the results from the study reported here showed a clearer pattern of alignment from proportional than from absolute measurements. Nevertheless, the results have shown that the comparison of the absolute and proportional measurements can be useful, e.g. the effect of the tail on the peak alignment was visible in both absolute and proportional measurements, whereas quantity-dependent differences only occured in the proportional measurements.

This study has shown that the effect of quantity on the peak alignment of VQ-words was still visible despite of the peak delay caused by the long tail context. Asu et al. (2009) showed that the early turning point in VQ3- words and the late turning point in VQ2- and VQ3-words were stable in phrase-initial, -internal and -final words of spontanous speech. The new finding of the present study is that the effect of quantity on the peak alignment of VQ-words was quite robust independently of the post-focal context (long vs. short tail contexts). Nevertheless, the results for the quantity-dependent peak alignment of VQ2-words in the current study differ from the results of previous studies (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). The present results have instead shown

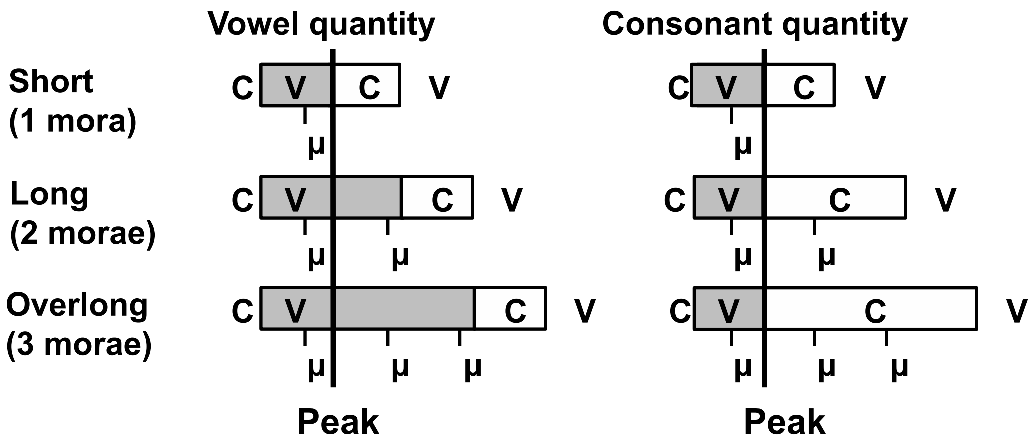


Figure 2.10: The location of the peak in relation to the offset of the first mora in VQ- and CQ-words.

the peak of VQ2-words to be nearer the temporal midpoint, rather than at the offset of the vowel. This is consistent with the findings of Krull (1997), who reported a peak at 55% of the vowel duration. A reason for the earlier alignment could be that e.g. Asu et al. (2009) and Lippus et al. (2013) used the so called turning point of the pitch contour, defined as the point in the pitch contour "where there is a noticeable change in the direction of the F0 contour in $V_1$ from rising or level to falling " (Asu et al., 2009, 52) whereas the current study and Krull (1997) used the peak. The peak is obviously earlier than the turning point.

The absolute and proportional peak alignment of VQ- and CQ-words was similar in the current study. The peak of VQ- and CQ-words was realised after a similar duration after the $V_1$ onset; differences occured only in the VQ2-, VQ3- vs. CQ2-, CQ3-words of the short tail context. These two cases excluded, the current study obtained a similar result for the absolute peak alignment as Lippus et al. (2013). However, the present study did not only find similarities in the absolute peak alignment of VQ- and CQ-words, but also in the peak alignment proportional to the $V_1C_2$-duration. In both VQ- and CQ-words the proportional peak alignment had the order Q3 $<$ Q2 $<$ Q1, where $<$ denotes that the timing of Q3-peaks was proportionally earlier than the timing of Q2-peaks and so on. Previous studies only reported peak alignment differences between VQ1-, VQ2- and VQ3-words (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013), but not between CQ1-, CQ2- and CQ3-words (Krull, 1992; Lehiste, 1997b; Lippus et al., 2013). Although the current study found peak alignment differences between CQ1-, CQ2- and CQ3-words similar to those of VQ-words, these differences only emerged, if the whole $V_1C_2$-sequence was taken into account. There were no alignment differences between CQ1-, CQ2- and CQ3-words, if the unit of observation was only $V_1$, because the peak of CQ-words was constantly aligned with the end of $V_1$, similarily as in the study of Lippus et al. (2013).

The development of the three quantity degrees of Estonian and especially the explanation for the difference between the long and overlong quantity degree is a much debated issue. According to e.g. Lehiste (2003), Veske (1879), Collinder (1929) and Hint (1980) the Estonian overlong quantity degree developed from the long quantity degree through compensatory lengthening due to aopocope and syncope, i.e. short vowels at the end or in the middle of a word were lost after long syllables. Ehala (2003) on the other hand provides a theory in which the long quantity degree emerged from the overlong quantity degree through shortening. There are at least two different phonological approaches for explaining the difference between the long and overlong quantity degree. The approach of Kager (1994) and Prince (1980) is a metrical one. The overlong quantity degree can build a foot by its own in monosyllabic words and this property of the overlong quantity degree is defining for the distinction between the long and overlong quantity degree in the concept of Kager (1994) and Prince (1980). Other authors tried to explain the difference between the long and overlong quantity degree with the help of morae (e.g. Bye, 1997; Eek and Meister, 1997; Ehala, 2003; Hayes, 1989; Hint, 1980). Traditionally the concept of the mora is used for the description of the Japanese language, which is said to be mora-timed (e.g Kubozono, 1996; Pierrehumbert and Beckman, 1988; Warner and Arai, 2001). For other languages under the term Moraic Theory (e.g. Hayes, 1989, 1995b; McCarthy and Prince, 1986) morae are used to e.g. explain syllable weight (Hayes, 1995b,a; McCarthy and Prince, 1986), compensatory lengthening (e.g. Hayes, 1989) and stress patterns (Hayes, 1995a). For Estonian the concept of the mora was used in the sense of Moraic Theory (e.g. Bye, 1997; Eek and Meister, 1997; Ehala, 2003; Hayes, 1989; Hint, 1980), but not in the sense of mora-timing (Lehiste, 1990). Estonian overlength is a challenge for Moraic Theory, because traditionally trimoraic syllables are unwelcome in Moraic Theory (Hayes, 1995b). Therefore the main point of discussion between the different moraic concepts of Estonian is the treatment of overlength. While e.g. Hayes (1989) allows trimoraic syllables for a couple of languages including Estonian, other authors try to avoid them for Estonian by using e.g. freestanding morae (Bye, 1997) or mora-splitting (Eek and Meister, 2003). Regardless of the discussion whether Estonian can have trimoraic syllables or not, the results of the current study created the impression that the alignment of the peak with the offset of the first mora could be the reason for the peak alignment differences in VQ- and CQ-words (Figure 2.10). For the visualisation (Figure 2.10) the trimoraic approach of Hayes (1989) was chosen. The results of the present experiment raise the possibility that the concept of the mora could be used as an anchorpoint for the peak within the $V_1C_2$-sequence (Figure 2.10). The results of the present study have shown the offset of the first mora in VQ1- and CQ-words to be at the $V_1C_2$-boundary and the location of the peak corresponded to this (Figures 2.2 and 2.3). In VQ2- and VQ3-words the offset of the first mora lies somewhere in $V_1$. Its location could only be estimated via a comparison with the peak alignment of CQ2- and CQ3-words. The peak alignment of VQ2- and CQ2-words and VQ3- and CQ3-words respectively was similar (Figures 2.2, 2.3, 2.4, and 2.5) and hence there is evidence that it could also correspond to the end of the first mora.

Anyhow, since the concept of the mora is not unproblematic for Estonian and especially the detection of the mora boundary in VQ2- and VQ3-words, a more simple interpretation of the results concerning the peak alignment similarities of VQ- and CQ-words should also be considered: firstly, in both VQ- and CQ-words the peak alignment in proportion to the $V_1C_2$-sequence is cued by the order Q3 $<$ Q2 $<$ Q1; secondly, the duration between the $V_1$-onset and the peak is roughly similar between VQ1- vs. CQ1-, VQ2- vs. CQ2- and VQ3- vs. CQ3-words; thirdly, the anchorpoint for the peak is the offset of $V_1$ in VQ1- and CQ1-words, a bit before the middle of the $V_1C_2$-sequence in VQ2- and CQ2-words and in the first third of VQ3- and CQ3-words.

## 2.5 Summary

This chapter analysed the influence of quantity and the number of post-focal unstressed syllables on the peak alignment of nuclear H*+L pitch accents in Estonian for both vowel (VQ) and consonant quantity (CQ). Previous literature regarding peak alignment differences due to quantity for falling accents did not report systematic differences for CQ-words (Krull, 1992; Lehiste, 1997b; Lippus et al., 2013), but only for VQ-words (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1997a, 2003; Liiv, 1961). The unit of observation in the studies cited was the vowel of the first stressed syllable of a word. The current study showed that if not only the vowel ($V_1$) but also the following consonant ($C_2$) was taken into account the peak alignment of VQ- and CQ-words was similar: In the short quantity degree the peak was located near the end of $V_1$. In the long quantity degree it was placed a bit before the middle of the $V_1C_2$-sequence. In the overlong quantity degree the peak was found in the first third of the $V_1C_2$-sequence. The results created the impression that the peak could be aligned with the offset of the first mora in both VQ- and CQ-words. Regarding the influence of the number of post-focal unstressed syllables the results of previous studies could also be confirmed. Similar to the results for non-quantity languages (Arvaniti and Ladd, 1995; Arvaniti et al., 1998; De la Mota, 2005; Möbius and Jilka, 2007; Mücke and Hermes, 2007; Peters, 1999; Prieto et al., 1995; Silverman and Pierrehumbert, 1990; Steele, 1986) it was found that the high target of H*+L accents in Estonian was shifted leftwards if the number of post-focal unstressed syllables decreased.

# Chapter 3

# Phrase-final lengthening

## 3.1 Introduction

The aim of this chapter is to analyse the domain of phrase-final lengthening (PFL) in Estonian vowel (VQ) and consonant (CQ) quantity words produced with an H*+L nuclear pitch accent. The results of the first chapter showed that PFL (also known under the term prepausal, preboundary or final lengthening) influenced the $V_1C_2$-duration of disyllabic sentence-final words of the form $C_1V_1C_2V_2$. However, the components of the $V_1C_2$-sequence of VQ- and CQ-words were affected differently (Figure 2.2 and 2.3). In VQ1-, VQ2-, VQ3- and CQ1- words it was mainly $V_1$ which was lengthened in sentence-final position (short tail words) whereas in CQ2- and CQ3-words $C_2$ was lengthened. The purpose of the current chapter is to investigate and to explain these differences further via an analysis of the domain of PFL in Estonian. The results will be discussed on the basis of three different models explaining PFL: a Structure-based, a Content-based and a Hybrid view model (Turk and Shattuck-Hufnagel, 2007).

So far PFL in Estonian has been found to affect the word (Lehiste, 1981; Mihkla, 2005, 2006) and the segment level (Asu et al., 2009; Eek and Meister, 2003; Krull, 1997) in both read (Eek and Meister, 2003; Lehiste, 1981; Mihkla, 2005, 2006) and sponantous speech (Asu et al., 2009; Krull, 1997). PFL in Estonian has only been investigated for VQ-words, but not for CQ-words. Furthermore, with the exception of Eek and Meister (2003), the studies analysing the segment level only concentrated on vowels (Asu et al., 2009; Krull, 1997). The influence of PFL on the vowel of the second syllable ($V_2$) of VQ-words seemed to be quite stable (Asu et al., 2009; Eek and Meister, 2003; Krull, 1997) while the duration of the vowel of the first syllable ($V_1$) was not always affected by the phrase-final position of the target word (Eek and Meister, 2003; Krull, 1997). While Asu et al. (2009) found an influence of PFL on $V_1$ and $V_2$ of VQ-words in all three quantity degrees for pitch accented and deaccented words, Krull (1997) reported a lengthening of both $V_1$ and $V_2$ only for VQ3-words. In VQ2-words only the duration of $V_2$ increased, and in VQ1-words it depended on the speaker whether either both vowels or only one of the vowels was lengthened by PFL. Eek and Meister (2003) investigated the vowels and the consonants of VQ-words. They found a longer duration of $C_2$ in sentence-final words and a lengthening of $V_2$ for all three quantitiy degrees. There was no lengthening of $V_1$ in VQ1-words. Nevertheless, the diphthong of the first syllable of VQ2- and VQ3-words differing in the quantity of this diphtong (/lauta/ (VQ2) vs. /lau:ta/ (VQ3)) was lengthened in phrase-final position. There is evidence of progressive lengthening in the data of Eek and Meister (2003), i.e. that the nearer a segment was to the final boundary, the more it was lengthened.

PFL has also been reported for Finnish (Hakokari et al., 2005; Nakai et al., 2009) and Hungarian (e.g. White and Mády, 2008), two other Finno-Ugric languages both having a two-way contrast of vowel and consonant quantity. As for Estonian, only VQ-words have been investigated so far. In Finnish PFL was found to operate at both word (Hakokari et al., 2005) and segmental levels (Nakai et al., 2009). Nakai et al. (2009) analysed disyllabic words differing in the quantity of the vowel of the first syllable ($C_1V_1C_2V_2$ vs. $C_1V_1V_1C_2V_2$). They found an influence of PFL on both the vowels and $C_2$ and in case of VQ1-words also on $C_1$. Furthermore they reported progressive lengthening for VQ2-words. In Hungarian, PFL was found for both VQ1- and VQ2-vowels in absolute final syllables and for VQ2-vowels also in penultimate syllables (White and Mády, 2008).

In previous studies, three major classes of models for explaining PFL were used: a Structure-based, a Content-based and a Hybrid view model (Turk and Shattuck-Hufnagel, 2007). In the Structure-based model, the domain of PFL is a fixed part of speech defined by linguistic structure, e.g. the final-syllable rhyme (Turk and Shattuck-Hufnagel, 2007; Wightman et al., 1992). In the Content-based view, PFL operates independently of the linguistic structure (Turk and Shattuck-Hufnagel, 2007). Those stretches of speech are lengthened by PFL which overlap with a so-called lengthening gesture such as the $\pi$-gesture (Byrd and Saltzman, 2003; Byrd et al., 2006). The

$\pi$-gesture is an abstract prosodic gesture which is anchored with its middle to the end of the final syllable. It overlaps with the segmental gestures, and during this overlap all articulatory movements are slowed down, hence the segments are lengthened (Byrd et al., 2006). The Hybrid view (Turk and Shattuck-Hufnagel, 2007) rests on the assumption that PFL lengthens the segments of a structurally fixed domain such as the rhyme of the final syllable. Nevertheless, if due to phonological quantity contrasts the rhyme cannot be lengthened unlimitedly, PFL can also affect earlier parts of speech (Cambier-Langeveld, 1997 for Dutch).

In American English PFL affected only syllable rhymes (Turk and Shattuck-Hufnagel, 2007; Wightman et al., 1992), which could be explained with a Structure-based model. In the study of Turk and Shattuck-Hufnagel (2007), PFL lengthened both the rhymes of the final and the main-stressed syllables whereas in Wightman et al. (1992) only the rhyme of the final syllable was affected. Turk and Shattuck-Hufnagel (2007) additionally found progressive lengthening: the rhyme of the final syllable was lengthened more than the rhyme of the main-stressed syllable.

Two hypotheses were tested in this chapter: firstly, whether - and consistently with a Structure-based view (Turk and Shattuck-Hufnagel, 2007) - the domain of PFL in Estonian VQ- and CQ-words is the rhyme of the stressed and the final syllable; secondly, whether - similarily as in American English (Turk and Shattuck-Hufnagel, 2007) and an earlier study of Estonian (Eek and Meister, 2003) - there is progressive lengthening in Estonian, i.e. the nearer a segment is to the final boundary the more it is lengthened.

## 3.2 Method

### 3.2.1 Subjects and materials

The same material from the same subjects as in the first chapter was analysed, i.e. there were four target word triplets of vowel and four target word triplets of consonant quantity (Table 2.1 and 2.2) spoken by nine speakers of Standard Estonian. The target words were embedded in two carrier phrases (Table 2.3), one carrier phrase with no unstressed syllables after the target word (final context) and one carrier phrase with two unstressed syllables after the target word (non-final context).

### 3.2.2 Analysis

The same utterances used for the study of the first chapter were also used for the current chapter, i.e. only those utterances produced with a nuclear H*+L pitch accent on the target word have been analysed. In total there were 103 utterances of vowel and 99 utterances of consonant quantity.

The utterances had already been divided into phonetic segments for the analysis of the first chapter. The same segmentation was also used for the current investigation.

The percentage lengthening of all sentence-final words ($PFL_{perc}$) was calculated separately for each segment and quantity degree from

$$PFL_{perc} = 100 \left[(DurF_{mean} - DurNF_{mean}) / DurF_{mean}\right]$$

where $DurF_{mean}$ is the mean duration of the appropriate segment of all sentence-final words (e.g. $C_1$ of all final VQ1-words) and $DurNF_{mean}$ the mean duration of the corresponding sgment of all non-final words in miliseconds.

For all analyses the R-software combined with the EMU/R-package (Harrington, 2010) was used. The statistical analysis was carried out using mixed models in R. The significance values were obtained by comparing a mixed model with and without the appropriate independent factor.

## 3.3 Results

Two hypotheses were tested: firstly, whether the domain of PFL in Estonian are the rhymes of the first and the second syllable and secondly, whether progressive lengthening occurs in Estonian.

PFL was found on both word and segment level of VQ- and CQ-words (Figures 3.1 and 3.2). In sentence-final position almost all segments of VQ- and CQ-words were lengthened, but to a different degree (Figures 3.3 and 3.4 and Tables 3.1 and 3.2). Exceptions were $C_1$ of VQ1- and CQ1-words and $V_1$ of CQ2-words. In these cases the duration did not change at all or was even reduced.

Mixed models were used to assess whether the segments were lengthened significantly. For each segment and each quantity degree a mixed model was carried out separately. The dependent variable was the segment duration.

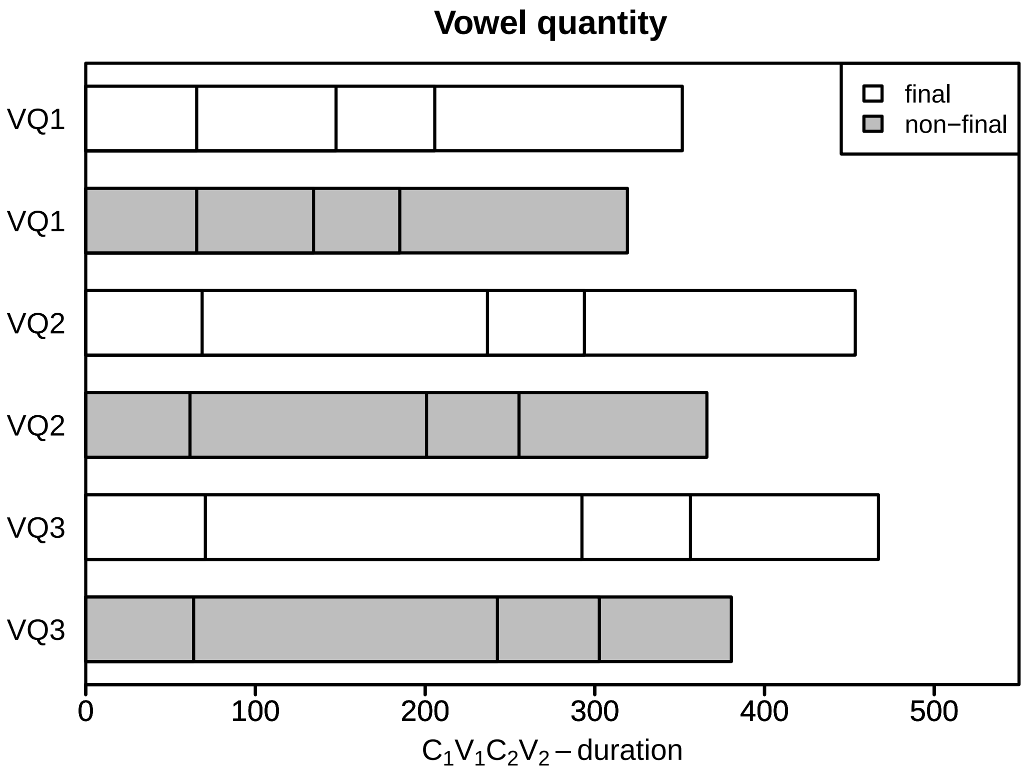


Figure 3.1: The $C_1V_1C_2V_2$-duration of VQ-words in final (white bars) and non-final (grey bars) position for the three quantity degrees.

Table 3.1: Growth of the segment duration from non-final to final VQ-words in percentage terms; significant growth in duration is printed in bold.

| **Segment** | **VQ1** | **VQ2** | **VQ3** |
|---|---|---|---|
| C1 | 0% | 10.45% | 9.8% |
| V1 | **16.15%** | **17.09%** | **19.33%** |
| C2 | **12.65%** | 4.64% | 6.07% |
| V2 | 8.03% | **30.63%** | **29.77%** |
| Word | 9.2% | **19.28%** | **18.55%** |

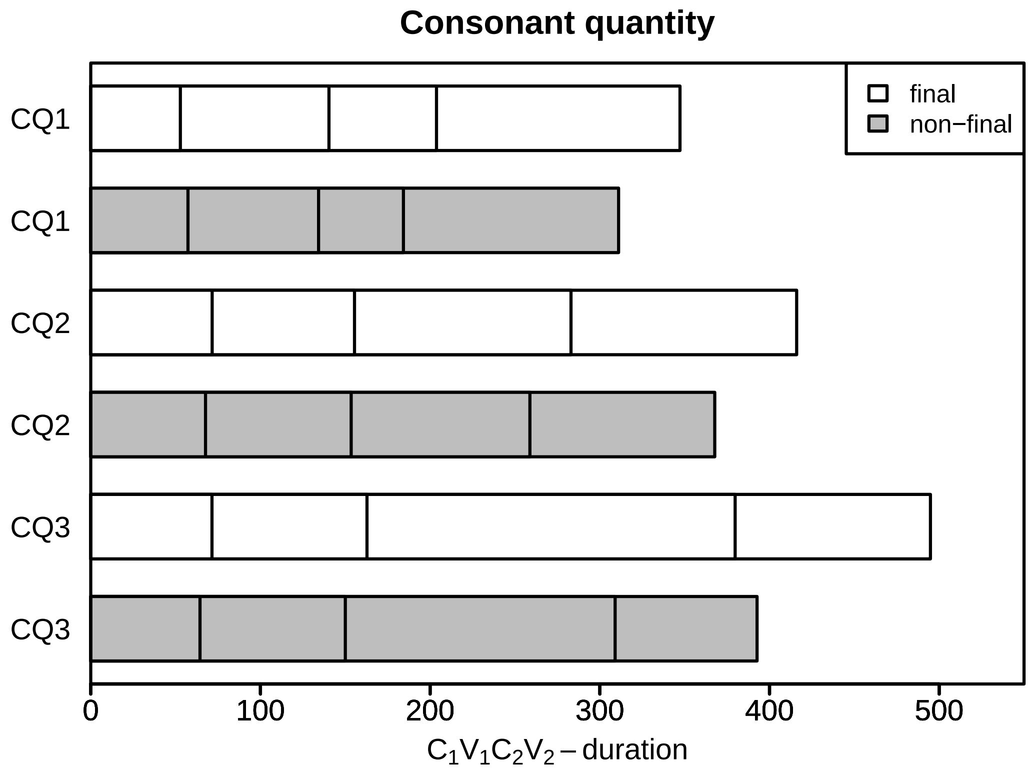


Figure 3.2: The $C_1V_1C_2V_2$-duration of CQ-words in final (white bars) and non-final (grey bars) position for the three quantity degrees.

Table 3.2: Growth of the segment duration from non-final to final CQ-words in percentage terms; significant growth in duration is printed in bold. Negative values indicate a decrease of segment duration.

| **Segment** | **CQ1** | **CQ2** | **CQ3** |
|---|---|---|---|
| C1 | -8.5% | 5.47% | **9.88%** |
| V1 | **12.11%** | -2.32% | 6.25% |
| C2 | **21.1%** | **17.44%** | **26.71%** |
| V2 | 11.63% | **18.08%** | **27.32%** |
| Word | **10.42%** | **11.6%** | **20.65%** |

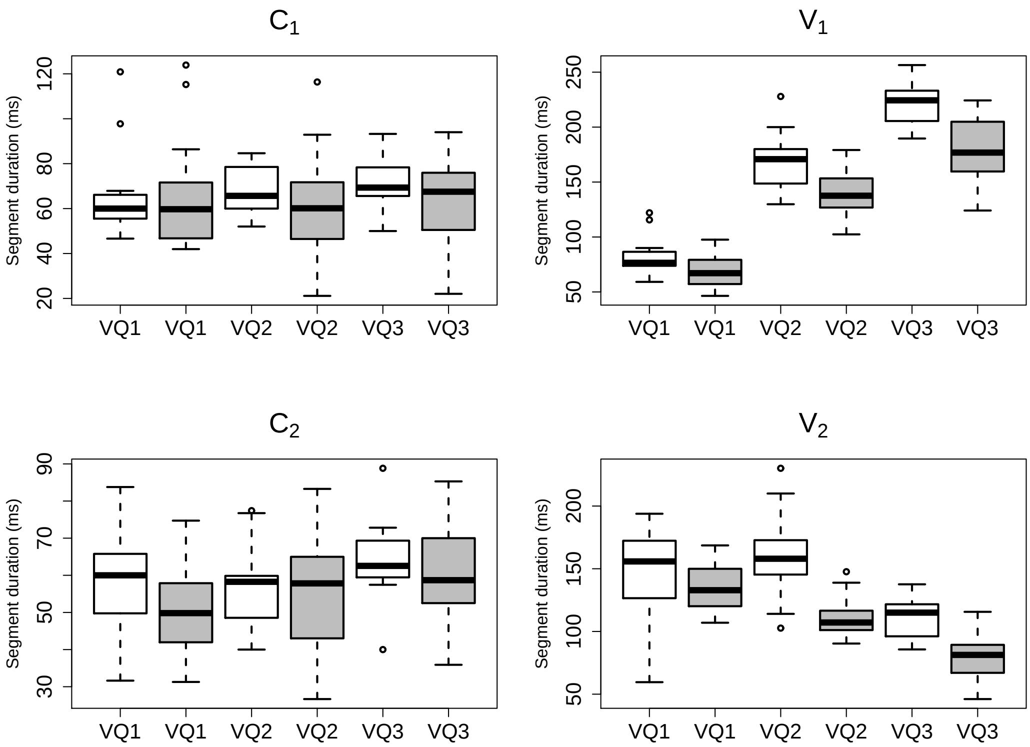


Figure 3.3: Segment durations of final (white boxes) and non-final (grey boxes) VQ-words separately for the three quantity degrees.

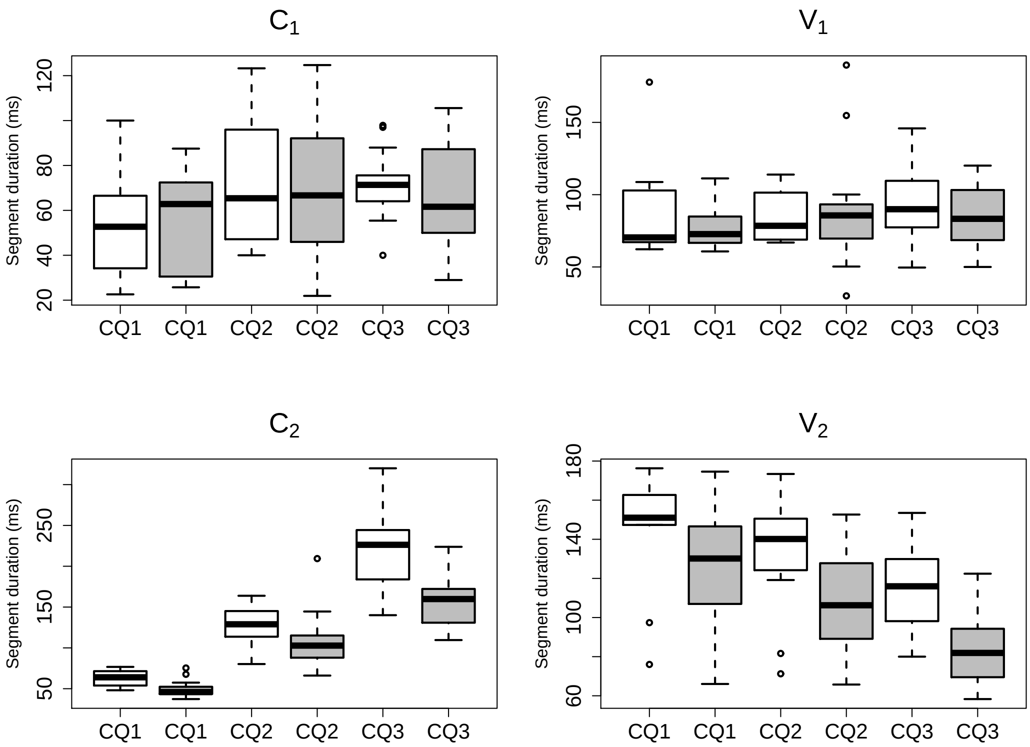


Figure 3.4: Segment durations of final (white boxes) and non-final (grey boxes) CQ-words separately for the three quantity degrees.

Table 3.3: The influence of Finality on the segment duration of VQ-words. Significant values of the random factor Word are printed in italics.

| **Segment** | **VQ1** | **VQ2** | **VQ3** |
|---|---|---|---|
| C1 | non-significant | non-significant | non-significant |
| V1 | $\chi^2[1] = 7.91$, $p < 0.01$ | $\chi^2[1] = 11.72$, $p < 0.001$ | $\chi^2[1] = 16.46$, $p < 0.001$ |
| C2 | $\chi^2[1] = 5.43$, $p < 0.05$ | non-significant | non-significant |
| | *$\chi^2[1] = 11.48$, $p < 0.001$* | *$\chi^2[1] = 8.84$, $p < 0.01$* | *$\chi^2[1] = 14.72$, $p < 0.001$* |
| V2 | non-significant | $\chi^2[1] = 26.43$, $p < 0.001$ | $\chi^2[1] = 18.45$, $p < 0.001$ |
| Word | non-significant | $\chi^2[1] = 25.87$, $p < 0.001$ | $\chi^2[1] = 18.76$, $p < 0.001$ |

Finality was a factor and Speaker and the Target Word were random factors. The results of the mixed models are listed in the Tables 3.3 and 3.4. In VQ1- and CQ1-words both $V_1$ and $C_2$ were lengthened significantly in sentence-final position. In VQ2- and VQ3-words $V_1$ and $V_2$ were affected signifantly by PFL, and in CQ2- and CQ3-words PFL influenced $C_2$ and $V_2$ significantly. In CQ3-words additionally $C_1$ was lenghtened significantly.

There was evidence for progressive lengthening in CQ1-, VQ2- and VQ3-words only (Tables 3.1 and 3.2). Those segments which were lengthened significantly were lengthened more the nearer they were to the sentence boundary. In CQ2- and CQ3-words $V_2$ was only slightly lengthened more than $C_2$ and was therefore not interpreted as progressive lengthening (Tables 3.1 and 3.2). In VQ1-words there was neither progressive lengthening, $C_2$ was lengthened less even though it was nearer to the boundary (Tables 3.1 and 3.2).

Table 3.4: The influence of Finality on the segment duration of CQ-words. Significant values of the random factor Word are printed in italics.

| **Segment** | **CQ1** | **CQ2** | **CQ3** |
|---|---|---|---|
| C1 | non-significant | non-significant<br>*$\chi^2[1] = 18.98$, $p < 0.001$* | $\chi^2[1] = 11.08$ $p < 0.001$<br>*$\chi^2[1] = 27.34$, $p < 0.001$* |
| V1 | $\chi^2[1] = 4.74$, $p < 0.05$<br>*$\chi^2[1] = 6.47$, $p < 0.05$* | non-significant<br>*$\chi^2[1] = 11.12$, $p < 0.001$* | non-significant<br>*$\chi^2[1] = 4.07$, $p < 0.05$* |
| C2 | $\chi^2[1] = 7.34$, $p < 0.01$<br>*$\chi^2[1] = 3.54$, $p < 0.1$* | $\chi^2[1] = 10.27$, $p < 0.01$ | $\chi^2[1] = 31.87$, $p < 0.001$ |
| V2 | non-significant | $\chi^2[1] = 13.11$, $p < 0.001$<br>*$\chi^2[1] = 4.33$, $p < 0.05$* | $\chi^2[1] = 19.22$, $p < 0.001$ |
| Word | $\chi^2[1] = 6.14$ $p < 0.05$ | $\chi^2[1] = 17.16$, $p < 0.001$<br>*$\chi^2[1] = 3.95$, $p < 0.05$* | $\chi^2[1] = 34.6$, $p < 0.001$ |

## 3.4 Discussion

The current chapter investigated the influence of PFL on VQ- and CQ-words in Estonian. It was found that in sentence-final position in VQ2- and VQ3-words $V_1$ and $V_2$ were lengthened significantly whereas in CQ2- and CQ3-words $C_2$ and $V_2$ were affected by PFL. $C_1$ was also found to be lengthened in CQ3-words. PFL influenced $V_1$ and $C_2$ in VQ1- and VQ2-words. Progressive lengthening was only found for VQ2-, VQ3- and CQ1-words.

The results of the current experiment differ to some extent from the results of previous studies investigating PFL in the three major Finno-Ugric languages Estonian, Finnish and Hungarian. Previous studies of Estonian (Asu et al., 2009; Eek and Meister, 2003; Krull, 1997), Finnish (Nakai et al., 2009) and Hungarian (White and Mády, 2008) reported a lengthening of $V_2$ independently of the quantity degree for VQ-words; however, in the current study $V_2$ was only lengthened in VQ2-, VQ3-, CQ2- and CQ3-words. The results of the influence of PFL on $V_1$ in the present study were consistent with those of Asu et al. (2009) for Estonian and Nakai et al. (2009) for Finnish who reported an influence of PFL on $V_1$ for all quantity degrees. However, the results of this experiment differ from those of Eek and Meister (2003) and Krull (1997) who found a quantity degree dependent $V_1$ lengthening. Nevertheless, the results of Eek and Meister (2003) for Estonian and Nakai et al. (2009) for Finnish regarding a lengthening of $C_2$ for VQ-words independently of the quantity degree were confirmed for VQ1-, CQ1-, CQ2- and CQ3-words.

A Content-based model, more precisely the $\pi$-gesture framework of Byrd et al. (2006), cannot explain all the results of the current study. The $\pi$-gesture is anchored with its middle to the end of the sentence-final word, and all segments of the segmental gesture overlapping with the $\pi$-gesture should be lengthened (Byrd et al., 2006). As far as Estonian is concerned, this means that segments between the offset of the word and somewhere within the word should be lengthened, since this is also the domain of the $\pi$-gesture. In CQ2-words the start of the $\pi$-gesture seems to lie somewhere during $C_2$ and extends until the end of $V_2$ since the adjacent segments $C_2$ and $V_2$ were affected by PFL. However, the results showed that PFL did not apply to the entire $C_1V_1C_2V_2$-sequence in VQ2- and VQ3-words, but only to $V_1$ and $V_2$ (but not $C_2$). This result is not compatible with the prediction from the $\pi$-gesture model.

According to the Hybrid view model, PFL should affect the final syllable rhyme and probably earlier stretches of speech, if a lengthening of the final-syllable rhyme is constrained (Cambier-Langeveld, 1997). In Estonian, the final-syllable rhyme cannot be lengthened unlimitedly, because Estonian quantities differ in the duration ratio of the first and the second syllable. A lengthening of only one of these syllables could destroy these characteristic ratios. The results of the current study confirmed the results of Cambier-Langeveld (1997) for Dutch by showing that in VQ2-, VQ3-, CQ2- and CQ3-words indeed not only the final-syllable rhyme was influenced by PFL, but also parts of the first syllable. Nevertheless, the Hybrid view is not plausible for a general explanation of PFL in Estonian since the final-syllable rhyme of VQ1- and CQ1-words was not lengthened at all.

The Structure-based model seems to be the best model for explaining the results of the current study. According to the Structure-based model, PFL should lengthen a fixed interval of speech defined by linguistic structure. However, the prediction according to this model that the rhyme of the stressed syllable and the final syllable should be lengthened was not supported by the results of the present experiment. There are two reasons for this: firstly, PFL did not affect the rhyme of the first syllable in VQ1- and CQ1-words. Secondly, it cannot be ensured that in CQ2- and CQ3-words indeed the rhyme of the first syllable and not only the onset consonant of the second syllable was lengthened since the syllable boundary lies somewhere during the intervocalic geminate $C_2$. However, the Structure-based model is still suitable for explaining PFL in Estonian if the linguistic structure is changed. There was one linguistic structure which was lengthened in all quantity degrees and quantity types: the main bearer of the quantity contrast, i.e. $V_1$ in VQ-words and $C_2$ in CQ-words.

Progressive lengthening, i.e. the nearer a segment to the end of the sentence the more it is lengthened, was found for Estonian VQ2-, VQ3- and CQ1-words similar as for American English (Turk and Shattuck-Hufnagel, 2007), Estonian VQ-words (Eek and Meister, 2003) and Finnish (Nakai et al., 2009). In CQ2- and CQ3-words, there was no progressive lengthening: C2 and V2 were lengthened about the same amount, probably because they were directly adjacent to each other. Maybe two adjacent segments in absolute final position were treated as one segment, and therefore the penultimate segment got the same amount of lengthening than the last segment. Adjacency alone without finality seems to be insufficient for an explanation, because in CQ1-words progressive lengthening was found for the adjacent segments $V_1$ and $C_2$ which were lengthened to a different degree. But an unresolved issue is why in constrast to CQ1-words, there was no progressive lengthening for VQ1-words: that is, why in VQ1-words, $V_1$ was lengthened to a greater extent than $C_2$. An investigation with more material could be useful. Furthermore it would be interesting to also study Q1-words, which can occur both as a member of a vowel quantity triplet (e.g. /vilu/ ('cool') - /viːlu/ ('slice', gen. sg.) - /viːːlu/ ('slice', part. sg.)) and as a member of a consonant quantity triplet (e.g. /vilu/ ('cool') - /villu/ ('Villu', gen. sg.) - /vilːlu/ ('wool', part. sg.)). Although this case is rare, its investigation would shed more light on the nature of the PFL domain in Estonian and on the generability of the results of the current study.

## 3.5 Summary

The purpose of this chapter was to analyse he domain of phrase-final lengthening (PFL) in Estonian for both vowel (VQ) and consonant quantity (CQ) words produced with a falling nuclear H*+L pitch accent. The results were interpreted in terms of three different models explaining PFL: a Structure-based, a Content-based, and a Hybrid view model. In earlier studies of Estonian, PFL had been found to influence both the word (Lehiste, 1981; Mihkla, 2005, 2006) and the segment level (Asu et al., 2009; Eek and Meister, 2003; Krull, 1997). However, with the exception of Eek and Meister (2003), the focus of the investigation in previous studies were the vowels, but not the consonants of disyllabic VQ-words. Moreover, CQ-words have not been studied at all so far. Evidence for PFL on the segment level also comes from two other Finno-Ugric languages: Hungarian (White and Mády, 2008) and Finnish (Nakai et al., 2009). For American English, PFL was found to influence syllable rhymes(Turk and Shattuck-Hufnagel, 2007; Wightman et al., 1992). In the present study, two hypotheses were tested: firstly, that PFL in Estonian only affects syllable rhymes and can be best explained with a Structure-based model; and secondly, that there is progressive lengthening in Estonian, i.e. the nearer a segment to the sentence boundary the more it is lengthened, as found for English (Turk and Shattuck-Hufnagel, 2007), Estonian VQ-words (Eek and Meister, 2003) and Finnish (Nakai et al., 2009). In the present study PFL influenced $V_1$ and $V_2$ in VQ2- and VQ3-words, and $C_2$ and $V_2$ in CQ2- and CQ3-words. $C_1$ was also found to be lengthened in CQ3-words. PFL affected $V_1$ and $C_2$ in VQ1- and CQ1-words. The Structure-based model was indeed the best model for explaining PFL in Estonian. Nevertheless, the site of the lengthening is unlikely to be the syllable rhyme. This is because in VQ1- and CQ1-words the rhyme of the first syllable was not lengthened at all while in CQ2- and CQ3-words it was unclear whether only the rhyme or/and the coda consonant of the first syllable were influenced by PFL. However, one linguistic structure was lengthened independently of the quantity degree and type: the main bearer of the quantity contrast, i.e. $V_1$ in VQ-words and $C_2$ in CQ-words. Other structures were also found to be lengthened, but depending on the quantity degree and type. Progressive lengthening was found for CQ1-, VQ2- and VQ3-words similar as in English (Turk and Shattuck-Hufnagel, 2007), an earlier study of Estonian (Eek and Meister, 2003) and Finnish (Nakai et al., 2009). Progressive lengthening was only present if the segments were not in adjacent final position as e.g. in CQ2- and CQ3-words. Nevertheless, it remaines unclear why there was no progressive lengthening in VQ1-words.

# Chapter 4

# Peak alignment and speaking rate

## 4.1 Introduction

The main result of chapter 2 was that their were similarities of the alignment of the starred tone of a nuclear H*+L pitch accent in Estonian in both VQ- and CQ-words. In the current chapter it is explored whether this alignment is stable with respect to a variation of the speaking rate.

Earlier studies of Estonian (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013) showed that VQ-words differ in their peak alignment, namely that in disyllabic words of the form $C_1V_1C_2V_2$ the peak of VQ3-words occurs early in $V_1$ whereas in VQ1- and VQ2-words the peak is located towards the end of the vowel. These alignment differences were only present in H*+L pitch accents (Asu et al., 2009). For CQ-words no comparable peak alignment differences were found (Krull, 1992; Lehiste, 1997b; Lippus et al., 2013). Nevertheless, a very recent study (Lippus et al., 2013) of spontanous speech compared the peak alignment of disyllabic VQ- and CQ-words produced with an H*+L pitch accent and found that the turning point of the high target of both CQ- and VQ-words was realised in a similar duration from the onset of $V_1$. The quantity type (VQ vs. CQ) did not have an influence on the location of the turning point. Also in chapter 2 similarities between the peak alignment of VQ- and CQ-words were found, namely that the peak alignment in proportion to the $V_1C_2$-sequence had the order $Q3 < Q2 < Q1$ in both VQ- and CQ-words, where $<$ denotes that the peak of Q3 was timed proportionally earlier than the peak of Q2 and so on. Furthermore the duration between the $V_1$ onset and the peak was similar between VQ- and CQ-words. The anchorpoint of the peak in VQ1- and CQ1-words was the offset of $V_1$, in VQ2- and CQ2-words a point a bit before the middle of the $V_1C_2$-sequence and in VQ3- and CQ3-words a point within the first third of the $V_1C_2$-sequence.

Chapter 2 showed that the aligment of the peak was sensitive to the number of post-focal unstressed syllables. If there were no unstressed syllables following the target word, the peak was timed to occur earlier in the vowel. The purpose of the current experiment is to investigate whether a variation of the speaking rate influences the alignment of the peak with the anchorpoints mentioned above.

For Estonian there are no studies investigating the influence of speaking rate on the tonal alignment of pitch accents yet. Only the influence of speaking rate on segment duration and on the quantity-dependent duration ratios between the first and the second syllable of disyllabic words have been studied so far (Eek and Meister, 2003). Nevertheless, there are several studies exploring the influence of speaking rate on the alignment of starred and unstarred tones of bitonal pitch accents in other languages (e.g. Caspers and van Heuven, 1993 for Dutch; Igarashi, 2004 for Russian; Ladd et al., 1999, Silverman and Pierrehumbert, 1990 for English; Nitisaroj, 2006 for Thai; Prieto and Torreira, 2007, De la Mota, 2005 for Spanish; Xu (1998) for Mandarin Chinese). Some of the studies cited found that the starred and unstarred tone of a bitonal pitch accent was anchored independently with the segmental string, and a change of speaking rate did not have an influence on neither the beginning, nor the end of the investigated rising pitch accent (Igarashi, 2004 for Russian; Ladd et al., 1999 for English and Xu, 1998 for Mandarin Chinese). Also in Thai speaking rate did not influence the alignment of peaks and valleys of the tones (Nitisaroj, 2006).

Nevertheless, there are also studies which found a strict anchoring only for the beginning of a rising pitch accent, but not for the end, i.e. speaking rate had an influence on the end of the rise, but not on the beginning (Caspers and van Heuven, 1993 for Dutch; De la Mota, 2005 for Castilian Spanish; Prieto and Torreira, 2007 for Peninsular Spanish). In Spanish rising accents the beginning of the rise was constant while the peak was earlier under a slower speech rate (De la Mota, 2005; Prieto and Torreira, 2007). Caspers and van Heuven (1993) found that the onset of a Dutch rising pitch accent was invariant under different speaking rates, but the end of the rise varied. Silverman and Pierrehumbert (1990) analysed the alignment of the peak, but not the beginning of the rise of English H* accents, and found a peak delay if speech rate decreased, i.e. the peak shifted in the opposite direction as in Spanish (Prieto and Torreira, 2007; De la Mota, 2005). Most of the studies cited mainly investigated the leading and the starred tone of a rising pitch accent (Caspers and van Heuven, 1993; De la Mota, 2005; Igarashi, 2004; Prieto and Torreira, 2007; Xu, 1998), only Caspers and van Heuven (1993) also investigated trailing tones and found that the beginning and the end of Dutch falls varied unsystematically under different speaking rates. They did not find any constant alignment.

To test whether speaking rate has an influence on the alignment of the starred tone of a falling H*+L pitch accent in Estonian, an experiment was designed varying the speaking rate (normal vs. fast), but controlling those factors which are known to influence the alignment of high and/or low tones such as e.g. the number of post-focal unstressed syllables (e.g. Arvaniti et al., 1998; De la Mota, 2005; Möbius and Jilka, 2007; Mücke and Hermes, 2007; Peters, 1999; Rathcke and Harrington, 2007; Silverman and Pierrehumbert, 1990; Steele, 1986), phonological vowel duration (e.g. Ladd et al., 2000; Plüschke, 2011), position of the word accent (e.g. Arvaniti et al., 1998; Prieto et al., 1995) and syllable structure (e.g. D'Imperio, 2000; Hellmuth, 2005, 2006; Ishihara, 2003; Prieto and Torreira, 2007; Welby and Lœvenbruck, 2005, 2006). It is assumed that speaking rate does not influence the alignment of the high target of H*+L pitch accents in Estonian, because besides segment duration peak alignment is used to differentiate between the quantity degrees (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013).

Two hypotheses were tested in the current chapter:

H1) The peak alignment of both VQ- and CQ-words is similar, i.e. in VQ1- and CQ1-words the peak is aligned with the offset of $V_1$, in VQ2- and CQ2-words a bit before the middle of the $V_1C_2$-sequence and in VQ3- and CQ3-words within the first third of the $V_1C_2$-sequence.

H2) Speaking rate does not have an influence on the peak alignment proportional to the $V_1C_2$-duration.

## 4.2 Method

### 4.2.1 Subjects and materials

Six female subjects (age range 20-55) participated in the experiment. One of the subjects had already participated in the experiment of chapter 2. All subjects were native speakers of Standard Estonian. Four of the subjects were paid for their participation. All recordings were conducted in Estonia, either in a quiet surrounding at the universities of Tartu and Tallinn or at the subjects' homes. A high quality headset was used for the recordings.

Eleven word triplets of vowel quantity (Table 4.1) and six word triplets of consonant quantity (Table 4.2) were chosen as target words. The target words contained only voiced segments. All target words were disyllabic and had the form $C_1V_1C_2V_2$. The triplets with vowel quantity differed only in the duration of $V_1$ and the consonant quantity triplets only in the duration of $C_2$. The eleven vowel quantity target word triplets were recorded twice and the six consonant quantity triplets four times.

The target words were embedded in carrier sentences in narrow focus position. The carrier sentences were designed to elicit a nuclear H*+L pitch accent on the target word. Because of differences in the grammatical category of the target word triplets, two different carrier sentences had to be constructed: one for the short and long quantity degree and another for the overlong quantity degree (Table 4.3). For the recordings the carrier sentences were presented as answers to questions to make it easier for the subjects to produce the target words in the narrow focus position.

Five subjects were recorded by the author with the help of the recording software *SpeechRecorder* (Draxler and Jänsch, 2004), which prompted first the question and then the corresponding answer (carrier sentence) on the screen of a notebook computer. Only the carrier sentences were recorded.The carrier sentences were presented in randomised order. The sixth subject did the recodings alone using her own computer and *Wikispeech* (Draxler and Jänsch, 2008), the online version of *SpeechRecorder*. The recordings were divided in two parts: at first the subjects read the carrier sentences in their normal speaking rate, and then they were asked to read the sentences as fast as possible. In total a number of 276 sentences was recorded for each subject.

Table 4.1: Target words with vowel quantity.

| **VQ1 (short)** (words in gen. and nom. sg.) | **VQ2 (long)** (words in gen. sg.) | **VQ3 (overlong)** (words in part. sg.) |
|---|---|---|
| /løma/ ('mash') | /løːma/ ('fray') | /løːːma/ ('fray') |
| /lori/ ('gossip') | /loːri/ ('veil') | loːːri ('veil') |
| /mæra/ ('mare') | /mæːra/ ('degree') | /mæːːra/ ('degree') |
| /mimi/ ('Mimi') | /miːmi/ ('mime') | miːːmi/ ('mime') |
| /mini/ ('very small, miniskirt') | /miːni/ ('mine') | /miːːni/ ('mine') |
| /mɤna/ ('spell') | /mɤːna/ ('low tide') | /mɤːːna/ ('low tide') |
| /rivi/ ('line') | /riːvi/ ('grater') | /riːːvi/ ('grater') |
| /vɑlu/ ('pain') | /vɑːlu/ ('swath') | /vɑːːlu/ ('swath') |
| /vɑrɑ/ ('estate') | /vɑːrɑ/ ('little hill') | /vɑːːrɑ/ ('little hill') |
| /vɑri/ ('something crumbled') | /vɑːri/ ('old man') | /vɑːːri/ ('old man') |
| /vilu/ ('cool') | /viːlu/ ('slice') | /viːːlu/ ('slice') |

Table 4.2: Target words with consonant quantity.

| **CQ1 (short)** (words in gen. and nom. sg.) | **CQ2 (long)** (words in gen. sg.) | **CQ3 (overlong)** (words in part. sg.) |
|---|---|---|
| /vilu/ ('cool') | /villu/ ('Villu') | /vilːlu/ ('wool', only part. pl.) |
| /vinɑ/ ('smell') | /vinnɑ/ ('hoist') | /vinːnɑ/ ('hoist') |
| /vɤlɑ/ ('debt') | /vɤllɑ/ ('gallows') | /vɤlːlɑ/ ('gallows') |
| /nɑri/ ('bunk bed') | /nɑrri/ ('fool') | /nɑrːri/ ('fool') |
| /lyli/ ('link') | /lylli/ ('gallows') | /lylːli/ ('gallows') |
| /linɑ/ ('linen') | /linnɑ/ ('city') | /linːnɑ/ ('city') |

Table 4.3: Carrier sentences for eliciting H*+L pitch accents on the target word.

| **Quantity degree** | **Carrier sentence** |
|---|---|
| Short/Long | *Me leiame* ____ *ju.* (We - find - ____ , too.) |
| Overlong | *Me nõuame* ____ *ju.* (We - demand - ____ , too.) |

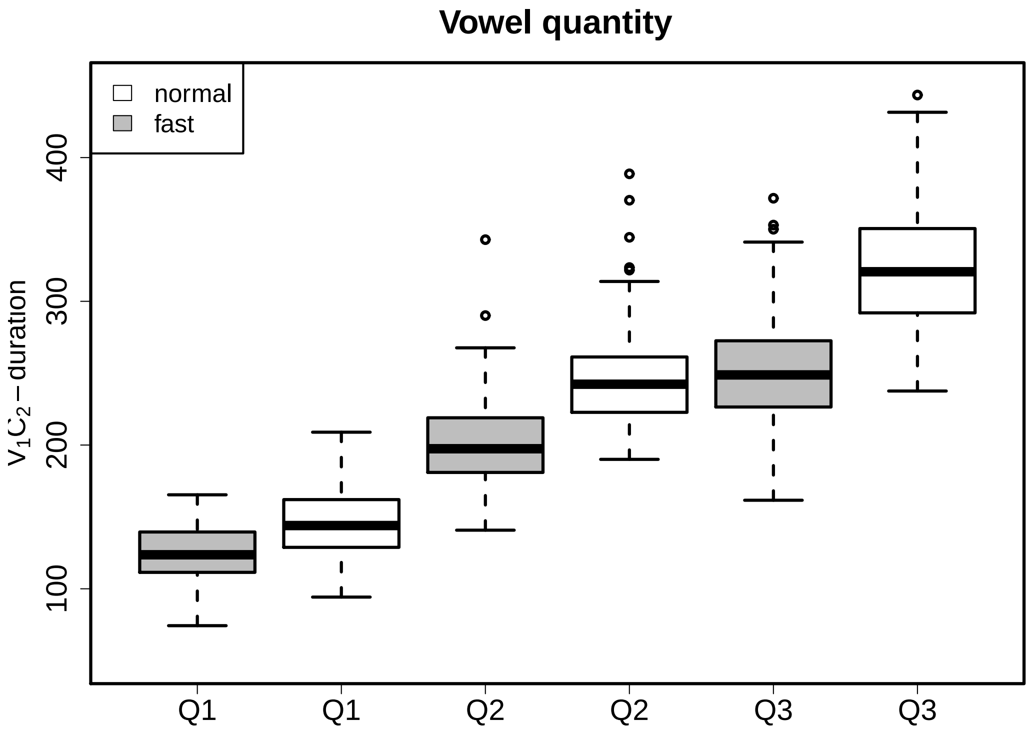


Figure 4.1: The $V_1C_2$-duration shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the normal (white boxes) and fast (grey boxes) speaking rate contexts for vowel quantity.

### 4.2.2 Analyis

The division into phonetic segments was done automatically with the help of the Munich Automatic Segmentation System MAUS (Schiel, 1999) and corrected manually with the *Praat* software (Boersma and Weenink, 2010). The F0 contours were determined by means of the Schaefer-Vincent periodicity detector (Schaefer-Vincent, 1983) in Emu (Harrington, 2010). All 1656 utterances were controlled for reading errors and the realised pitch accent. After excluding the utterances realised with an H+L* pitch accent (7.8%), containing reading errors (1.5%) or a deaccented target word (2.2%), a total of 1465 (88.5%) utterances produced with an H*+L pitch accent remained for the analysis.

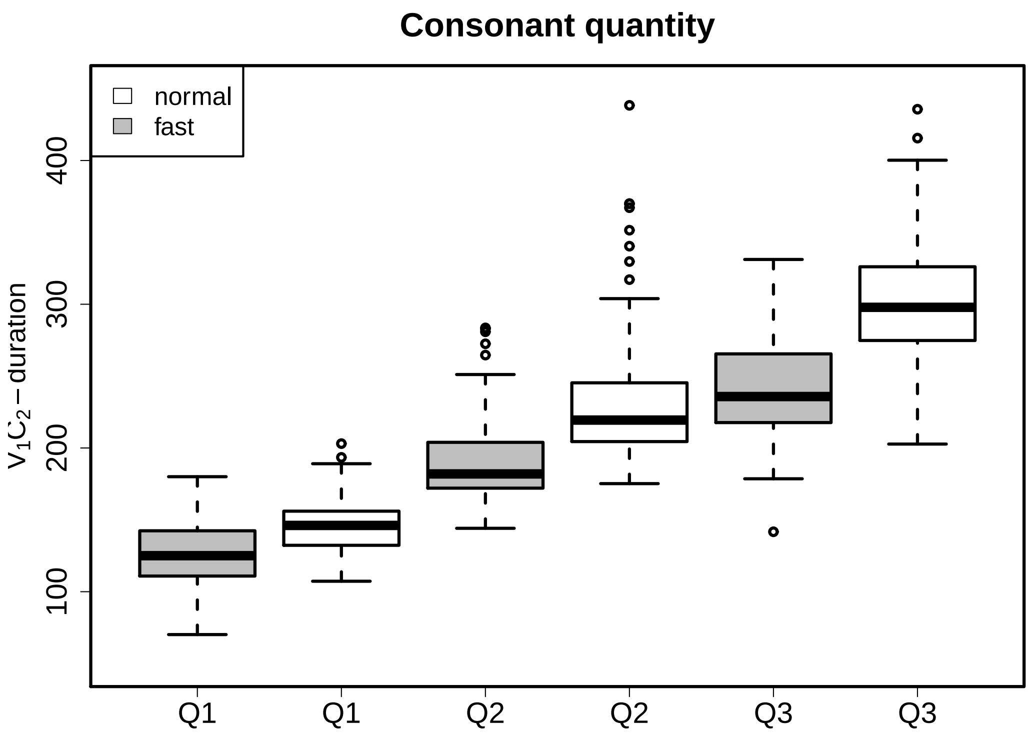


Figure 4.2: The $V_1C_2$-duration shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree combined with the normal (white boxes) and fast (grey boxes) speaking rate contexts for consonant quantity.

The subjects' ability to produce speech at two rates was checked prior to analysing the remaining utterances further. Because the peak will be later expressed in proportion to the duration of $V_1$ and $C_2$, the $V_1C_2$-duration of normal and fast utterances was compared with each other. The Figures 4.1 and 4.2 show that the speaking rate variation was successful in all three quantities and in both vowel and consonant quantity.

To confirm this observation, an RM-ANOVA was done with the dependent variable $V_1C_2$-duration and the independent variables Quantity degree (three levels: Q1, Q2, Q3), Speaking rate (two levels: normal, fast) and Quantity Type (two levels: vowel and consonant quantity). Indeed Speaking rate had a significant influence on the $V_1C_2$-duration ($F[1,5] = 91.5$, $p < 0.001$), and there was a significant interaction between Speaking rate and Quantity ($F[2,10] = 45.1$, $p < 0.001$) and between Speaking rate and Quantity Type ($F[1,5] = 44.4$, $p < 0.001$). However, post hoc tests confirmed that the difference between normal and fast speaking rate was significant for all three quantity degrees (Q1: $padj < 0.01$; Q2: $padj < 0.05$; Q3: $padj < 0.01$) and both quantity types (consonant quantity: $padj < 0.01$; vowel quantity: $padj < 0.01$).

To prepare those utterances produced with an H*+L accent for the analysis, the F0 contour of the target word and the first segment of the following word (/j/ of 'ju') were smoothed with the help of a discrete-cosine-transformation (DCT), similar as in the experiment of chapter 2. Before the smoothing, pitch-tracking errors were deleted and approximated to the neighbouring F0 values. A pilot study was made to detect the optimal number of coefficients for the smoothing. A smoothing with less than six coefficients was found to smooth the contour too much, therefore the smoothing was done with six coefficients.

After the smoothing, the values of the F0 maximum H* and the L[1] of the H*+L pitch accent were detected automatically, and the contours were checked manually whether the smoothing and the detection of the H* and the L was succesful. After excluding the errors of the smoothing and the H* and

[1]The L was detected with the help of zerocrossings of the first derivations of the smoothed signals, because the zerocrossings of the first derivation correspond to maxima and minima in the original signal. If there were no zerocrossings in the first derivation, the zerocrossings of the second derivations were used. Those signals were excluded where the second derivation neither had zerocrossings. The L detection is not described in more detail here, because the L will not be used in the analyses of this chapter. Nevertheless, the L detection was already performed for further purposes.

L detection, a number of 1193 (83.4%) of the 1465 utterances remained for further analyses. The remaining 1193 utterances consisted of 572 utterances of vowel quantity and 621 of consonant quantity.

The peak of the smoothed signal was expressed in absolute and proportional terms. The absolute peak alignment, $T_{abs}$ was calculated from

(1) $T_{abs} = T - V_{on}$

where T is the time of the H* of the H*+L pitch accent and $V_{on}$ the acoustic onset of the vowel of the first syllable (V1) of the target word.

The proportional peak alignment $T_{VCprop}$ was calculated relative to the acoustic onset of V1 and offset of C2 from

(2) $T_{VCprop} = T_{abs}/(C_{off} - V_{on})$

where $V_{on}$ is the acoustic onset of V1 and $C_{off}$ is the acoustic offset of C2.

All analyses were carried out with the R-software combined with the EMU/R-package (Harrington, 2010). RM-ANOVAs and mixed models in R were used for the statistical analyses. For the mixed models significant values were obtained by comparing a mixed model with and without the appropriate independent factor.

## 4.3 Results

Two hypotheses were tested. Firstly, whether the peak of both vowel (VQ) and consonant quantity (CQ) is similar and secondly, whether the alignment of the peak is invariant under changes of speaking rate, i.e. speaking rate should not have an influence on the proportional peak alignment $T_{VCprop}$.

Compatibly with H1, Figures 4.4 and 4.3 suggest the peak alignment is similar in both VQ- and CQ-words. In VQ1- and CQ1-words the peak was aligned with the offset of $V_1$, in VQ2- and CQ2-words a bit before the middle of the $V_1C_2$-sequence and in VQ3- and CQ3-words in the first third of the $V_1C_2$-sequence. Figures 4.5 and 4.6 also suggest consistently with H1 that the proportional peak alignment $T_{VCprop}$ was similar for both VQ- and CQ-words and had the order Q3 < Q2 < Q1, where < denotes that the timing of Q3-peaks was proportionally earlier than that of Q2-peaks and so on (Figures 4.5 and 4.6).

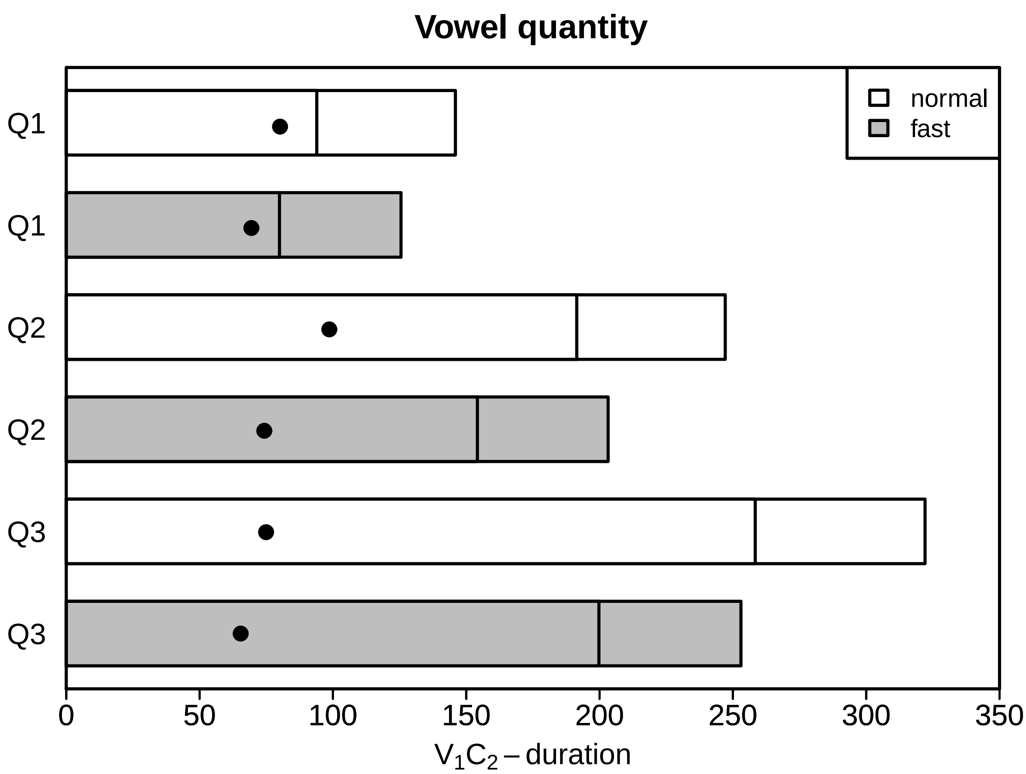


Figure 4.3: The bars show the $V_1C_2$-duration separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree for normal (white bars) and fast (grey bars) speaking rate for VQ-words. The lines indicate the border between $V_1$ and $C_2$. The dots mark the mean value of $T_{abs}$.

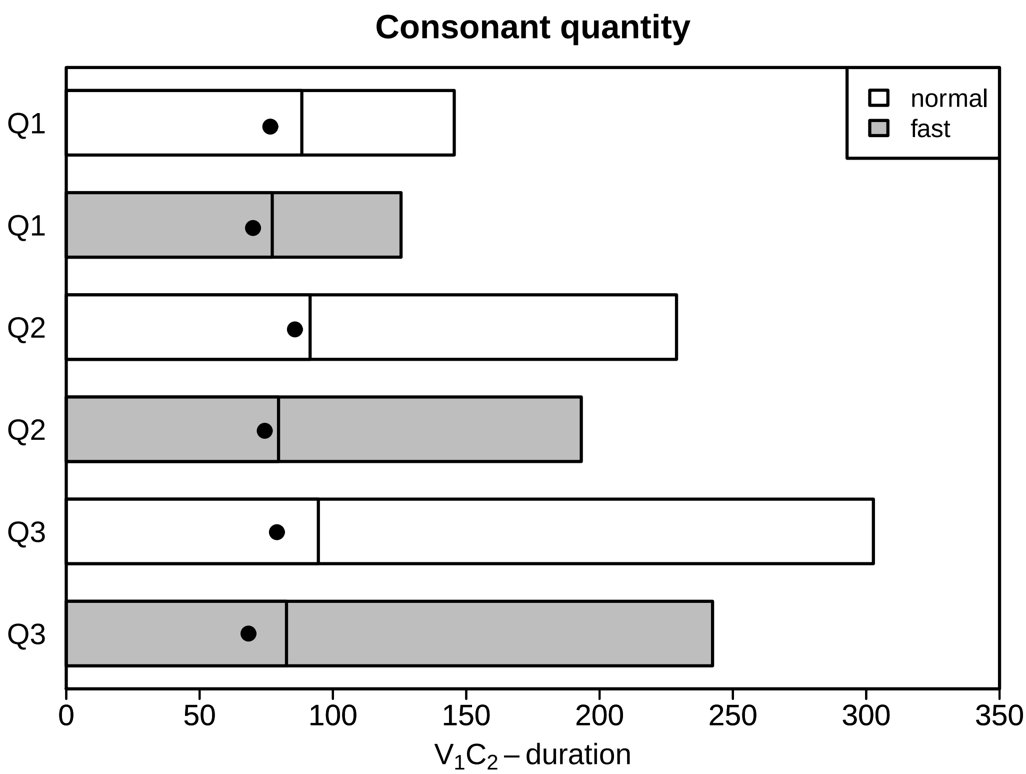


Figure 4.4: The bars show the $V_1C_2$-duration separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree for normal (white bars) and fast (grey bars) speaking rate for CQ-words. The lines indicate the border between $V_1$ and $C_2$. The dots mark the mean value of $T_{abs}$.

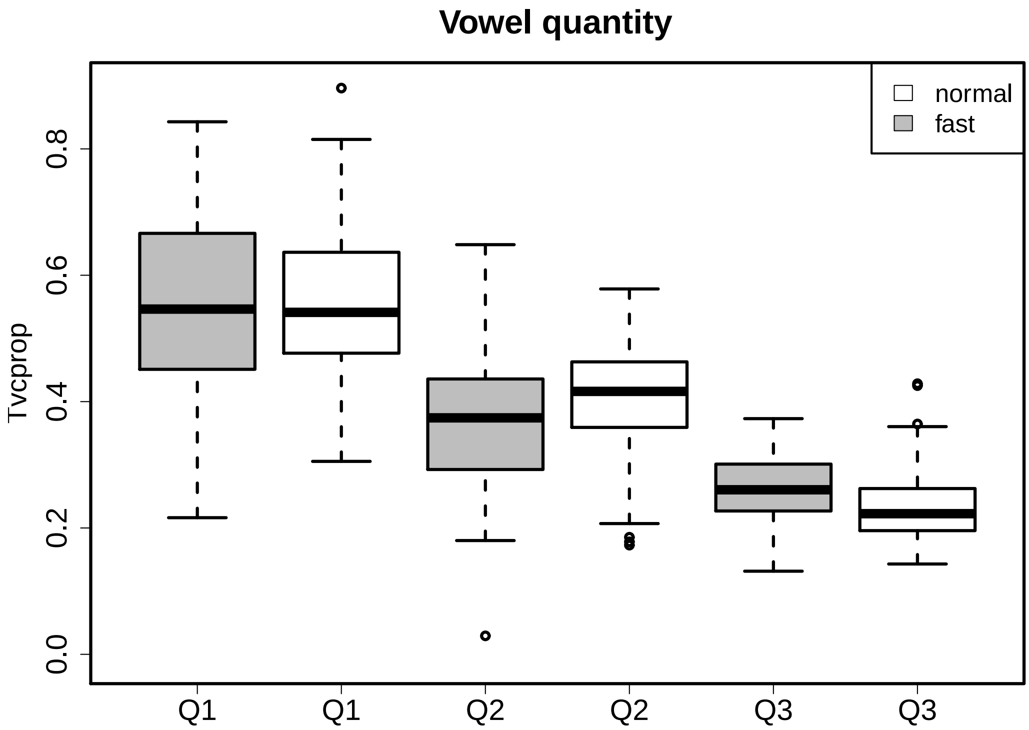


Figure 4.5: The proportional time for the tone target, $T_{VCprop}$, shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree for normal (white boxes) and fast (grey boxes) speaking rate for VQ-words.

According to H2, speaking rate should not have an influence on the alignment of $T_{VCprop}$, and indeed, there was hardly any influence of speaking rate on the alignment of $T_{VCprop}$ in neither VQ-words (Figure 4.5) nor CQ-words (Figure 4.6) independently of the quantity degree.

The observations made above regarding the alignment of $T_{VCprop}$ were statistically confirmed with the help of an RM-ANOVA and a mixed model. Firstly, an RM-ANOVA was done with the dependent variable $T_{VCprop}$ and the independent variables Speaking Rate (two levels: normal, fast), Quantity degree (three levels: Q1, Q2, Q3) and Quantity type (two levels: VQ, CQ). Speaking rate and Quantity type did not have a significant influence on the alignment of $T_{VCprop}$ whereas Quantity degree did ($F[2,10] = 102.8$, $p < 0.001$). Furthermore, there was a significant interaction between Speaking

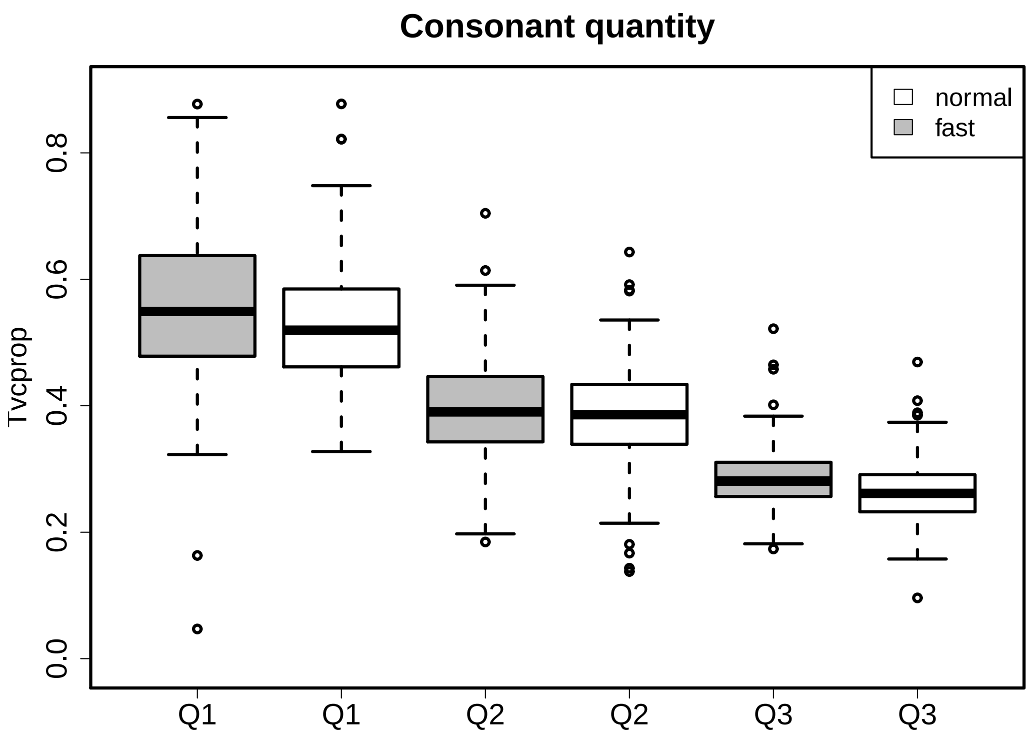


Figure 4.6: The proportional time for the tone target, $T_{VCprop}$, shown separately for the short (Q1), long (Q2) and overlong (Q3) quantity degree for normal (white boxes) and fast (grey boxes) speaking rate for CQ-words.

rate and Quantity Type ($F[1,5] = 21.7$, $p < 0.01$). Nevertheless, a post-hoc test for the different combinations of Speaking Rate and Quantity type did not show any significant combinations.

A mixed model with the dependent variable $T_{VCprop}$, the factors Speaking rate, Quantity degree and Quantity type and the random factors Speaker and Target word confirmed the results of the RM-ANOVA. Also in the mixed model, there was a significant interaction between Quantity Type and Speaking Rate ($\chi^2[1] = 4.3$, $p < 0.05$), but again post-hoc tests did not show any significant combinations of the factors Speaking Rate and Quantity type. However, the post-hoc test showed that all quantity degrees differed from each other significantly in the alignment of $T_{VCprop}$ in both normal speaking rate (VQ1vs. VQ2: $z$ = -11.7, $p <$ 0.001; VQ2 vs. VQ3: $z$ = -13.8, $p <$ 0.001; VQ1 vs. VQ3: $z$ = 25.9, $p <$ 0.001; CQ1 vs. CQ2: $z$ = -12.5, $p <$ 0.001; CQ2 vs. CQ3: $z$ = -9.5, $p <$ 0.001; CQ1 vs. CQ3: $z$ = -22.2, p $<$ 0.001) and fast speaking rate (VQ1 vs. VQ2: $z$ = -13.7, $p <$ 0.001; VQ2 vs. VQ3: $z$ = -8.7, $p <$ 0.001; VQ1 vs. VQ3: $z$ = -22.1, $p <$ 0.001; CQ1 vs. CQ2: $z$ = -12.5, $p <$ 0.001; CQ2 vs. CQ3: $z$ = -8.7, $p <$ 0.001; CQ1 vs. CQ3: $z$ = -21, $p <$ 0.001).

## 4.4 Discussion

There were two main findings of the current experiment: Firstly, that the peak alignment of VQ- and CQ-words was similar, if the unit of observation was the $V_1C_2$-sequence, and secondly, that this alignment was stable in respect to a variation of the speaking rate (normal vs. fast speaking rate).

The current experiment confirmed the results of earlier studies on Estonian reporting a peak early in the vowel for VQ3-words and late in the vowel for VQ1-words (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). Nevertheless, the results for VQ2-words of the current study differed from the results of the studies cited. Whereas the above-mentioned studies also reported a peak at the end of the vowel or in the second half of the vowel for VQ2-words, in the current study the VQ2-peak was instead found in the middle of the vowel, similar as in Krull (1997) and in chapter 2. As already mentionend in chapter 2, a reason for the earlier alignment could be that e.g. Asu et al. (2009) and Lippus et al. (2013) used the turning point and the current study the peak, which is obviously a bit earlier than the turning point. It remains unclear whether the location of the turning point of the current experiment would also be at the end of the vowel as in Asu et al. (2009) and Lippus et al. (2013) or still earlier. Nevertheless, the results of the present experiment confirmed the findings of chapter 2, which found a similar peak alignment of VQ- and CQ-words. As in chapter 2, the results of the current study created the impression that the peak could be aligned with the offset of the first mora (Figure 2.10). Since the speech material of the two studies was similar, a study with other and more material could be useful.

A comparison of the peak alignment of the target words followed by two unstressed syllables of chapter 2 and the target words of the current experiment, which were followed by one unstressed syllable, showed a similar proportional and absolute peak alignment. Whereas in chapter 2 it was shown that the peak alignment was sensitive to the number of post-focal unstressed syllables (the peak was timed earlier if there were no unstressed syllables following the target word), the current study showed that the alignment of the peak was invariant with regard to a variation of the speaking rate.

One of the main findings of the current study was that the peak alignment of VQ- and CQ-words was similar. Lippus et al. (2013) found similarities in the peak alignment of disyllabic VQ- and CQ-words extracted from spontanous speech. In the study of Lippus et al. (2013) the turning point of the high target of H*+L pitch accents was realised after a similar duration from the $V_1$ onset in both VQ- and CQ-words. This result is in line with the findings of the current study; a slight difference only occured in VQ2- vs. CQ2-words in the normal speaking rate context. Lippus et al. (2013) further found that in CQ1- and CQ2-words the turning point was located slightly after the offset of V1 and in CQ3-words slightly before it. In the present study, the peak of CQ-words was always located slightly before the V1 offset. A reason for these differences could be the use of the turning point by Lippus et al. (2013) and of the peak by the current study, which is a bit earlier than the turning point. Another reason could be that Lippus et al. (2013) analysed spontanous speech whereas the current experiment used read speech. Nevertheless, despite this slight difference the results of Lippus et al. (2013) for CQ-words are similar to the findings for CQ-words of the current experiment.

In the current study speaking rate did not have an influence on the proportional alignment of the peak, i.e. the peak was consistently aligned with its anchorpoint regardless of the speaking rate. This result is in line with earlier studies for other languages which neither found an influence of speaking rate on the alignment of starred tones of rising accents (Igarashi, 2004 for Russian; Ladd et al., 1999 for English and Xu, 1998 for Mandarin Chinese). Furthermore the results of the current study are similar to those of Nitisaroj (2006) for Thai pitch peaks, which neither were influenced by a variation of the speaking rate. Nevertheless, the results of the present experiment differ from the results of those studies, which found an influence of the speaking rate on the starred tones of rising pitch accents (Silverman and Pierrehumbert, 1990 for English; Caspers and van Heuven, 1993 for Dutch; Prieto and Torreira, 2007 and De la Mota, 2005 for Spanish). The results of the current study also differ from the results of Caspers and van Heuven (1993) for falling pitch accents in Dutch. Whereas Caspers and van Heuven (1993) did not find any consistent anchor points of neither the beginning nor the end of a fall, the present study did so for the high target of Estonian falling H*+L pitch accents. The results of the current experiment regarding a stable alignment of the peak independently of the speaking rate favour an interpretation of the peak alignment in Estonian in the sense of the segmental anchoring hypothesis (Ladd et al., 1999), i.e. that tonal targets are independently anchored with the segmental string. Nevertheless, the results of the present study only concern the starred tone of an Estonian H*+L pitch accent; a further investigation is needed to determine whether the L targets are also aligned independently with the segmental string or whether they are realised in respect to the starred tone.

## 4.5 Summary

The aim of this chapter was to test whether the high target of Estonian H*+L pitch accents is consistently aligned regardless of a variation of the speaking rate. The influence of speaking rate on peak alignment in Estonian had not been previously studied. Nevertheless, there are contradictory results from other languages. In some studies speaking rate did not have an influence on the peak alignment (Igarashi, 2004 for Russian; Ladd et al., 1999 for English; Nitisaroj, 2006 for Thai and Xu, 1998 for Mandarin Chinese) while in other studies speaking rate indeed influenced the peak alignment (Caspers and van Heuven, 1993 for Dutch; Prieto and Torreira, 2007 and De la Mota, 2005 for Spanish; Silverman and Pierrehumbert, 1990 for English). The results of the current study showed that the peak of both VQ- and CQ-words was aligned similarily, if the unit of observation was the $V_1C_2$-sequence. Speaking rate did not have an influence on the proportional peak alignment. The results are in line with an earlier study of Estonian which found that the absolute peak alignment of VQ- and CQ-words was similar (Lippus et al., 2013) and studies which found that the three quantity degrees differed in the peak alignment (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). Regarding speaking rate, the results of the current study favour an interpretation in the sense of the segmental anchoring hypothesis (Ladd et al., 1999) and are similar to those studies which did not find an influence of speaking rate on the peak alignment (Igarashi, 2004; Ladd et al., 1999; Nitisaroj, 2006; Xu, 1998).

# Chapter 5

# Summary and conclusion

The aim of this dissertation was threefold: firstly, to investigate how peak alignment is influenced by quantity; secondly to determine the way this relationship is further influenced by other prosodic factors such as the proximity to sentence boundaries and speaking rate; thirdly to analyse the influence of the sentence boundary on segment durations and its interaction with quantity. This was done via an analysis of Estonian which does not only have a three-way quantity contrast (short, long and overlong) but also two quantity types (vowel and consonant quantity).

Previous studies of Estonian found that words participating in the three-way quantity contrast differed in their peak alignment: the peak of short and long vowel quantity (VQ) words was located towards the end of the vowel of the stressed syllable whereas the peak of overlong VQ-words occured early in the vowel (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). Regarding these quantity-dependent peak alignment differences, three different research questions were adressed in this dissertation: firstly, whether the peak alignment differences of VQ-words found by previous studies are anchored toa certain point in the segmental string; secondly, whether consonant quantity (CQ) words show similar peak alignment differences as VQ-words; and thirdly, whether the quantity-dependent peak alignment differences are stable in regard to two prosodic factors: the vicinity of the sentence boundary and a variation of the speaking rate.

Earlier studies showed that segment durations of Estonian VQ-words were influenced by the vicinity to the sentence boundary, i.e. phrase-final lengthening (PFL) was found (e.g. Krull, 1997; Asu et al., 2009). Nevertheless, so far for Estonian neither differences between the influence of PFL in regard to the quantity type were explored nor has there been any attempt made to explain PFL with a phonological model. This dissertation adressed these two questions by investigating the interaction of PFL with the quantity degree (short, long vs. overlong) and quantity type (VQ- vs. CQ-words) and by trying to find the most suitable phonological model for explaining PFL in Estonian.

The main results of the three experiments of this dissertation are summarised below. Since experiment one (chapter 2) and experiment three (chapter 4) adressed the peak alignment, they are both presented in section 5.1. The results regarding PFL in Estonian are discussed in section 5.2. A general discussion (section 5.3) and an outlook (section 5.4) completes this chapter.

## 5.1 Peak alignment differences in Estonian

Previous studies on Estonian showed that VQ-words in Estonian differed in the realisation of the pitch contour: while short (Q1) and long (Q2) disyllabic VQ-words of the form $C_1V_1C_2V_2$ had a late peak in $V_1$, the peak of overlong (Q3) VQ-words was located early in $V_1$. In VQ3-words the fall of the pitch was completed in the first syllable whereas in VQ1- and VQ2-words it continued into the second syllable (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). The alignment differences were only present in words produced with an H*+L pitch accent (Asu et al., 2009). For CQ-words, no comparable peak alignment differences were found (Krull, 1992; Lehiste, 1997b; Lippus et al., 2013).

Tonal alignment research found that the number of post-focal unstressed syllables influenced the peak alignment: in particular, the high target of a bitonal pitch accent was found to be aligned earlier in a word the fewer the number of unstressed syllables following the accented syllable (Arvaniti et al., 1998; Arvaniti and Ladd, 1995; De la Mota, 2005; Möbius and Jilka, 2007; Mücke and Hermes, 2007; Peters, 1999; Prieto et al., 1995; Rathcke and Harrington, 2007; Silverman and Pierrehumbert, 1990; Steele, 1986). Nevertheless, in the data of Dalton and Ní Chasaide (2005) this peak shift was not present.

In contrast to the number of post-focal unstressed syllables, speaking rate was shown to have no influence on the alignment of the beginning and the end of a rising pitch accent (Igarashi, 2004; Ladd et al., 1999; Xu, 1998).

Other studies (Caspers and van Heuven, 1993; De la Mota, 2005; Prieto and Torreira, 2007; Silverman and Pierrehumbert, 1990) found no influence of the speaking rate on the beginning of a rising pitch accent, but on the end. Caspers and van Heuven (1993) did not find any consistent influence of speaking rate on the alignment of the beginning and the end of a Dutch fall.

In this dissertation two different experiments were designed to test the influence of the number of post-focal unstressed syllables and the speaking rate on peak alignment in Estonian and the interaction of these prosodic factors with quantity. Since the quantity-dependent peak alignment differences only occur in disyllabic words produced with an H*+L pitch accent (Asu et al., 2009), in both experiments disyllabic target words of the form $C_1V_1C_2V_2$, differing in either the quantity of $V_1$ (VQ-words) or $C_2$ (CQ-words), were embedded in carrier sentences to elicit a nuclear H*+L pitch accent on the target word.

### 5.1.1 Variation of unstressed syllables

In the first experiment (chapter 2), the number of unstressed syllables following the target word was varied: the target word was either in sentence-final position (short tail context) or followed by two unstressed syllables (long tail context). It was hypothesised that in line with earlier studies on other languages (Arvaniti and Ladd, 1995, Arvaniti et al., 1998 for Greek; De la Mota, 2005, Prieto et al., 1995 for Spanish; Peters, 1999, Möbius and Jilka, 2007, Mücke and Hermes, 2007, Rathcke and Harrington, 2007 for German; Silverman and Pierrehumbert, 1990, Steele, 1986 for English), the high target of the H*+L pitch accent would be earlier in the short tail context. Nevertheless, the quantity-dependent peak alignment for VQ-words (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013) should be still maintained. For CQ-words, no peak alignment differences between the quantity degrees were assumed, similarily as in the studies of Krull (1992), Lehiste (1997b) and Lippus et al. (2013).

One of the main findings of the first experiment was that the number of post-focal unstressed syllables indeed had an influence on the peak alignment in Estonian in both absolute and proportional measures. Just as in non-quantity languages (Arvaniti and Ladd, 1995, Arvaniti et al., 1998 for Greek; De la Mota, 2005, Prieto et al., 1995 for Spanish; Peters, 1999, Möbius and Jilka, 2007, Mücke and Hermes, 2007, Rathcke and Harrington, 2007 for German; Silverman and Pierrehumbert, 1990, Steele, 1986 for English), the peak was earlier in the short tail context, presumably because of the time pressure caused by the sentence boundary.

Despite this variation of the peak alignment due to the number of unstressed syllables, VQ1-, VQ2- and VQ3-words clearly differed in their peak alignment. As in earlier studies (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013), the peak of VQ3-words occured early in the vowel and the peak of VQ1-words towards the offset of the vowel. But in contrast to other studies (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013), the results from the present dissertation did not find the peak of VQ2-words to be timed close to the offset of the vowel, but rather in the middle of the vowel, like in Krull (1997). A reason for the differences could be that the current study and Krull (1997) used the peak but Asu et al. (2009) and Lippus et al. (2013) the turning point which is obviously a bit later than the peak.

As in Krull (1992), Lehiste (1997b) and Lippus et al. (2013), there were no peak alignment differences in CQ1-, CQ2- and CQ3-words if the peak was expressed in proportion to the $V_1$-duration because it was consistently aligned with the $V_1$ offset, as in Lippus et al. (2013). But there were similar peak alignment differences for VQ- and CQ-words when the time of the peak was calculated as a proportion of the $V_1C_2$-duration. It was found that in both VQ- and CQ-words the peak alignment proportional to the $V_1C_2$-duration had the order $Q3 < Q2 < Q1$, where $<$ denotes that the peak of Q3-words was timed proportionally earlier than the peak of Q2-words and so on.

The results of the current study created the impression that the peak of VQ- and CQ-words in Estonian was consistently aligned with a phonological landmark, i.e. the offset of the first mora of the word (see Figure 2.10). In VQ1-, CQ1-, CQ2- and CQ3-words, the peak was aligned with the offset of $V_1$ which corresponds to the offset of the first mora. In VQ2- and VQ3-words, the mora offset lies somewhere during $V_1$, and its location could only be assumed via a comparison with the proportional peak alignment of CQ2- and CQ3-words. Since the proportional peak alignment of VQ2- and VQ3- vs. CQ2- and CQ3-words was similar, it could be assumed that also in VQ2- and VQ3-words the peak was aligned with the offset of the first mora. Nevertheless, since the concept of the mora is not unproblematic for Estonian and the mora boundary in VQ2- and VQ3-words lies somewhere within $V_1$, a more simple interpretation of the results should also be considered: In VQ1- and CQ1-words the peak is aligned with the offset of $V_1$, in VQ2- and CQ2-

words with a point a bit before the middle of the $V_1C_2$-sequence and in VQ3- and CQ3-words in the first third of the $V_1C_2$-sequence.

### 5.1.2 Variation of the speaking rate

In the third experiment (chapter 4), the stability of the peak alignment in regard to a variation of the speaking rate was tested. For this purpose, the target words were embedded in a carrier sentence in narrow focus position to elicit a nuclear H*+L pitch accent. The target words were followed by one unstressed syllable to avoid a peak shift due to the sentence boundary. The carrier sentences were read in normal and fast speaking rate. Two hypotheses were tested: firstly, the peak of both VQ- and CQ-words should be similar and secondly, since the peak in Estonian is used to differentiate between the quantity degrees, speaking rate should not have an influence on the alignment of the peak proportional to the $V_1C_2$-duration.

The results of the current experiment confirmed the first hypothesis: as in the first experiment (chapter 2), the peak alignment of VQ- and CQ-words was similar. Both the absolute and the proportional peak alignment was similar as in the first experiment, and again, the peak alignment in proportion to the $V_1C_2$-duration had the order Q3 $<$ Q2 $<$ Q1, where the peak of Q3 was located earlier than the peak of Q2 and so on. As in the first experiment, the location of the peak of VQ2-words differed from earlier studies on Estonian where the peak was rather found towards the end of $V_1$ (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). In the current study the peak was rather aligned with the middle of $V_1$, as in Krull (1997). The reason for this could lie in the use of the turning point by Asu et al. (2009) and Lippus et al. (2013) as opposed to the peak by the current study and Krull (1997).

The second hypothesis was also confirmed: speaking rate did not influence the peak alignment in proportion to the $V_1C_2$-duration. This result is in line with studies on other languages reporting that speaking rate did not influence the alignment of starred tones of rising accents (Igarashi, 2004 for Russian; Ladd et al., 1999 for English and Xu, 1998 for Mandarin Chinese) or Thai pitch peaks (Nitisaroj, 2006). Nevertheless, the results of the current experiment differ from those studies finding an influence of the speaking rate on the alignment of rising pitch accents (Silverman and Pierrehumbert, 1990 for English; Caspers and van Heuven, 1993 for Dutch; Prieto and Torreira, 2007, De la Mota, 2005 for Spanish) and on the beginning and end of falls in Dutch (Caspers and van Heuven, 1993).

## 5.2 Phrase-final lengthening in Estonian

The results of the first experiment (chapter 3) showed that the segment durations of $V_1$ and $C_2$ of a disyllabic word of the form $C_1V_1C_2V_2$ were influenced differently by sentence-finality depending on the quantity type (VQ- vs. CQ-words). Therefore it was assumed that phrase-final lengthening (PFL) in Estonian may affect VQ- and CQ-words differently. The purpose of the second experiment (chapter 3) of this dissertation was to investigate the domain of PFL in Estonian and the interaction of PFL with quantity degree and type. A study was also made of which phonological model was best able to account for PFL in Estonian.

Previous studies of PFL in Estonian mainly concentrated on the vowels of disyllabic VQ-words. A lengthening of $V_2$ was found to be quite stable across studies and quantity degrees while a lengthening of $V_1$ was rather unstable and varied with the quantity degree and the study (Asu et al., 2009; Eek and Meister, 2003; Krull, 1997). Eek and Meister (2003) additionally investigated the consonants of VQ-words and found a lengthening of $C_2$ in all quantity degrees.

Earlier studies of PFL on other languages used three different models to explain PFL: a Structure-based, a Content-based and a Hybrid-view model (Turk and Shattuck-Hufnagel, 2007). In the Structure-based model, the lengthened part of speech is defined by linguistic structure, e.g. the rhyme of the final and the stressed syllable as found for American English (Turk and Shattuck-Hufnagel, 2007). In the Content-based model, PFL lengthens a part of speech overlapping with an abstract prosodic lengthening gesture such as the $\pi$-gesture (Byrd et al., 2006). The Hybrid-view model assumes that PFL influences a structurally fixed domain, e.g. the rhyme of the final syllable; nevertheless, earlier stretches of speech can be affected if the rhyme cannot not be lenghtened unlimitedly due to factors such as phonological quantity contrasts (Cambier-Langeveld, 1997).

Previous studies on English (Turk and Shattuck-Hufnagel, 2007), Estonian VQ-words (Eek and Meister, 2003) and Finnish (Nakai et al., 2009) found progressive lengthening, i.e. the nearer a segment to the final boundary the more it is lengthened.

Two hypotheses were tested in the second experiment: firstly, that the domain of PFL in Estonian is the rhyme of the first and the second syllable of a disyllabic word (and can be best accounted for in terms of a Structure-based model); and secondly, that progressive lengthening occurs in Estonian VQ- and CQ- words.

The same data was used for the analysis as in the first experiment (chapter 2), i.e. the target words were either followed by two (non-final context) or none unstressed syllables (sentence-final context).

It was found that PFL interacted with the quantity degree and type. While in VQ1- and CQ1- words PFL affected $V_1$ and $C_2$ significantly, in VQ2- and VQ3-words $V_1$ and $V_2$ were lenghtened and in CQ2- and CQ3-words $C_2$ and $V_2$.

The hypothesis that the domain of PFL in Estonian is the rhyme of the first and the second syllable could only be confirmed for VQ2- and VQ3-words. The results of PFL could nevertheless be explained in terms of a Structure-based model. However, the linguistic structures influenced by PFL were not the syllable rhymes as in American English (Turk and Shattuck-Hufnagel, 2007), but the main bearer of the quantity contrast, i.e. $V_1$ in VQ-words and $C_2$ in CQ-words. Other segments were also lengthened depending on the quantity degree and type.

Progressive lengthening was also found in the present experiment. This is therefore consistent with results that have been found for American English(Turk and Shattuck-Hufnagel, 2007), an earlier study of Estonian VQ-words (Eek and Meister, 2003) and Finnish (Nakai et al., 2009). In the current study progressive lengthening was only found for VQ2-, VQ3- and CQ1-words, and it was assumed that progressive lengthening only occured if the lengthened segments were not in adjacent final position.

## 5.3 General conclusion

E.g. Lehiste (2003) claimed that a sound change and the emergence of Estonian from Finnish could explain the peak alignment differences in Estonian VQ2- and VQ3-words. However, the reasons for the peak alignment differences remain still unclear. The results of the current dissertation give some evidence that the peak alignment differences can be explained with the help of the segmental anchoring hypothesis (see e.g. Ladd et al., 1999, 2000; Schepman et al., 2006) claiming that tonal targets are aligned with segmental anchors in a consistent way. The segmental anchor for Estonian could probably be the offset of the first mora - independently of the quantity degree and type. Differences in the proportional peak alignment emerge because of the interaction of the segmental anchorpoint with the temporal correlates of the quantity contrast.

Both experiments of this dissertation concerning peak alignment induced a time pressure for the realisation of the pitch accent: either the time for realising the pitch accent was reduced because the number of post-focal unstressed syllables decreased; or a faster speaking rate led to shorter segment durations. However, peak alignment was influenced only by the first condition. The reason for this could lie in what Ladd (2008) called a phonological vs. phonetic definition of time pressure. He claimed that time pressure due to a phonological reason (e.g. the number of syllables between the pitch accent and a boundary or other pitch accents) would influence the peak alignment whereas time pressure induced by a phonetic reason (e.g. the actual duration between the pitch accent and the boundary or other pitch accent) would have no effect on the peak alignment. The results of the current dissertation seem to confirm this claim if the time pressure due to a variation of the number of post-focal unstressed syllables is regarded as phonological and the time pressure due to a faster speaking rate as phonetic.

The sentence boundary in Estonian did not only influence the peak alignment but also the segment durations. The influence of PFL on segment durations strongly depended on the quantity degree and type, but nevertheless one linguistic structure was affected in all quantity degrees and types: the main bearer of the quantity contrast, i.e. $V_1$ in VQ-words and $C_2$ in CQ-words. This result was interpreted in the sense of a Structure-based model for explaining PFL; however, the main bearer of the quantity contrast was not the only linguistic structure affected by PFL: depending on the quantity degree and type additionally other segments were lengthened. There is therefore no support for a type of Structure-based model that assumes only one linguistic structure. However, the Structure-based model still remains the most plausible model for explaining PFL in Estonian.

## 5.4 Outlook

This dissertation investigated only the alignment of the high target of an H*+L pitch accent. Further studies should also explore the alignment of the trailing tone L, more precisely whether the L is aligned independently with the segmental string or in respect to the starred tone, and how this alignment interacts with quantity degree and type. The aims of such an investigation are at least threefold: firstly, the role of the L in regard to the distinction of the quantity degrees and potentially to the quantity types in Estonian should be explored; secondly, clarification is needed on the relation between starred and unstarred tones within an autosegmental-metrical theory of intonational phonology; thirdly, a further test is needed of the strict segmental anchoring hypothesis of Ladd et al. (1999), i.e. whether both tonal targets of a bitonal pitch accent are independently aligned with the segmental string.

Nevertheless, further studies on Estonian pitch accents should not be restricted to an investigation of tonal alignment only. In fact, the scaling of tonal targets as well as the duration and slope of the pitch accents should also be investigated, since these parameters may also contribute to the perception of tonal categories (see Niebuhr, 2007 for German).

In the current dissertation only nuclear pitch accents have been studied. Future research should also include prenuclear accents and investigate e.g. whether the nuclear and prenuclear pitch accents differ in the tonal alignment and the scaling or the slope; and whether prosodic factors, such as the number of post-focal unstressed syllables or the speaking rate, influence prenuclear pitch accents differently than nuclear accents.

This dissertation investigated the influence of the sentence boundary on both the alignment of the peak and segment durations. Further studies should also explore the influence of final lengthening induced by minor prosodic boundaries on the peak alignment and segment durations.

Future research should also include perception experiments to validate the results from the production experiments. Firstly, further perception studies should explore the role of the peak alignment and the role of the offset of the first mora as a probable anchorpoint on the quantity perception of both VQ- and CQ-words. Secondly, perception studies should investigate whether the peak shift and the lengthening of certain segments found for sentence-final words are used by listeners as a perceptual cue for perceiving a sentence boundary.

# Appendix A

# Zusammenfassung der Dissertation in deutscher Sprache

Die vorliegende Dissertation verfolgt drei Ziele:

(1) eine Untersuchung quantitätsbedingter Unterschiede in der Synchronisierung von Intonationsgipfeln (*peak alignment*),

(2) eine Analyse der Interaktion dieser Synchronisierungsunterschiede mit zwei prosodischen Faktoren (herannahende Satzgrenze und Variation des Sprechtempos), sowie

(3) eine Untersuchung des Einflusses der Satzgrenze auf Segmentdauern (phrasenfinale Längung) und deren Interaktion mit Quantität.

Die Analysen wurden anhand des Estnischen durchgeführt, da es als Quantitätssprache ideal für eine solche Untersuchung ist. Das Estnische weist nicht nur drei Quantitätstufen (kurz - Q1, lang - Q2 und überlang - Q3) auf, sondern auch zwei Quantitätstypen: Vokal- (VQ) und Konsonantenquantität (KQ). Der dreifache Quantitätskontrast betrifft zweisilbige Wörter der Form $K_1V_1K_2V_2$. VQ-Wörter unterscheiden sich in der Quantität von $V_1$, wohingegen bei KQ-Wörtern $K_2$ betroffen ist.

Diese Dissertation besteht aus insgesamt drei Experimenten. Die Hauptergebnisse dieser Experimente werden im Folgenden zusammengefasst und diskutiert. Da sich sowohl das erste (Kapitel 2) als auch das dritte Experiment (Kapitel 4) mit Gipfelsynchronisierung beschäftigen, werden beide im Unterkapitel A.1 beschrieben. Unterkapitel A.2 enthält die Ergebnisse des Experiments über phrasenfinale Längung.

## A.1 Unterschiede in der Gipfelsynchronisierung des Estnischen

Frühere Studien zum Estnischen (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013) zeigten, dass sich zweisilbige Vokalquantitäts-Wörter (VQ-Wörter) der Form $K_1V_1K_2V_2$ in der Realisierung der Grundfrequenzkontur unterscheiden: Kurze (Q1) und lange (Q2) VQ-Wörter weisen einen Gipfel gegen Ende von $V_1$ auf, wohingegen der Gipfel von überlangen (Q3) VQ-Wörtern früh in $V_1$ auftritt. Der Grundfrequenzabfall nach dem Gipfel ist in VQ3-Wörtern bereits innerhalb der ersten Silbe beendet, in VQ1- und VQ2-Wörtern dauert er noch in die zweite Silbe hinein an. Die beschriebenen Grundfrequenzunterschiede betreffen nur Wörter, die mit einem H*+L Tonakzent (*pitch accent*) realisiert werden (Asu et al., 2009). Für KQ-Wörter wurden keine vergleichbaren Grundfrequenzunterschiede gefunden (Krull, 1992; Lehiste, 1997b; Lippus et al., 2013).

Studien zur tonalen Synchronisierung fanden, dass die Anzahl an unbetonten Silben, die auf eine akzentuierte Silbe folgen, einen Einfluss auf die Gipfelsynchronisierung hat: der gesternte Ton (*starred tone*) eines bitonalen Tonakzents wird umso früher im Wort realisiert, je weniger unbetonte Silben auf die akzentuierte Silbe des Zielwortes folgen (Arvaniti und Ladd, 1995 und Arvaniti et al. 1998 für das Griechische; De la Mota, 2005 und Prieto et al., 1995 für das Spanische; Peters, 1999, Möbius und Jilka, 2007, Mücke und Hermes, 2007 und Rathcke und Harrington, 2007 für das Deutsche; Silverman und Pierrehumbert, 1990 und Steele, 1986 für das Englische). Im Irischen wurde diese Gipfelverschiebung nicht beobachtet (Dalton and Ní Chasaide, 2005).

Einige Studien zur tonalen Synchronisierung zeigten, dass Sprechgeschwindigkeit keinen Einfluss auf den Beginn und das Ende eines steigenden Tonakzents hat (Igarashi, 2004 für das Russische; Ladd et al., 1999 für das Englische and Xu, 1998 für das Mandarinchinesische). Trotzdem gibt es auch Studien, die fanden, dass Sprechgeschwindigkeit zwar keinen Einfluss auf den Beginn eines steigenden Tonakzents hat, aber auf das Ende (Caspers und van Heuven, 1993; De la Mota, 2005; Prieto und Torreira, 2007). Caspers und van Heuven (1993) fanden keinen systematischen Einfluss der Sprechgeschwindigkeit auf den Beginn und das Ende fallender Tonakzente des Niederländischen.

Um den Einfluss der Anzahl an unbetonten Silben und der Sprechgeschwindigkeit auf die Gipfelsynchronisierung und deren Interaktion mit Quantität im Estnischen zu testen, wurden zwei Produktionsexperimente entworfen. In beiden Experimenten wurden zweisilbige Zielwörter ($K_1V_1K_2V_2$) verwendet, die sich nur in der Quantität von $V_1$ (VQ-Wörter) bzw. $K_2$ (KQ-Wörter) unterschieden. Die Zielwörter wurden in Trägersätze eingebettet und nur die Wörter wurden für die Analyse verwendet, die tatsächlich mit einem H*+L Tonakzent realisiert wurden.

### A.1.1 Variation der Anzahl an unbetonten Silben

Im ersten Experiment dieser Dissertation (Kapitel 2) wurde die Anzahl an unbetonten Silben nach dem Zielwort variiert. Auf das Zielwort folgten entweder zwei unbetonte Silben (langer Nachlauf) oder keine (kurzer Nachlauf). Es wurde angenommen, dass bei einem kurzen Nachlauf der gesternte Ton eines H*+L Tonakzents früher im Wort realisiert wird als bei einem langen Nachlauf. Trotz des Einflusses des Nachlaufs sollten die quantitätsbedingten Unterschiede in der Gipfelsynchronisierung jedoch noch vorhanden sein. Für KQ-Wörter wurde angenommen, dass es keine vergleichbaren Unterschiede in der Gipfelsynchronisierung gibt wie in VQ-Wörtern.

Eines der Hauptergebnisse des ersten Experiments ist, dass die Anzahl der unbetonten Silben tatsächlich einen Einfluss auf die absolute und proportionale Gipfelsynchronisierung hat. Ähnlich wie in früheren Studien über Nicht-Quantitätssprachen (Arvaniti und Ladd, 1995 und Arvaniti et al. 1998 für das Griechische; De la Mota, 2005 und Prieto et al., 1995 für das Spanische; Peters, 1999, Möbius und Jilka, 2007, Mücke und Hermes, 2007 und Rathcke und Harrington, 2007 für das Deutsche; Silverman und Pierrehumbert, 1990 und Steele, 1986 für das Englische) führt der kurze Nachlauf zu einem Zeitdruck (*time pressure*) und der Gipfel wird nach links verschoben.

Trotz dieser Linksverschiebung unterscheidet sich die Gipfelsynchronisierung von VQ1-, VQ2- und VQ3-Wörtern im aktuellen Experiment deutlich. Ähnlich wie in früheren Studien zum Estnischen (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013) ist der Gipfel von VQ3-Wörtern früh und von VQ1-Wörtern spät im Vokal synchronisiert. Der Gipfel von VQ2-Wörtern ist jedoch eher in der Mitte des Vokals zu finden wie in Krull (1997) und nicht gegen Ende des Vokals wie in einer Reihe anderer Studien (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). Ein Grund könnte sein, dass im aktuellen Experiment und in Krull (1997) der Gipfel verwendet wird, in den Studien von Asu et al. (2009) und Lippus et al. (2013) hingegen der Wendepunkt (*turning point*) der F0-Bewegung. Der Wendepunkt ist von Natur aus etwas später als der Gipfel synchronisiert.

Ebenso wie in früheren Studien zum Estnischen (Krull, 1997; Lehiste, 1997b; Lippus et al., 2013) finden sich in der aktuellen Studie keine Unterschiede in der Gipfelsynchronisierung von KQ-Wörtern, wenn der Gipfel in Proportion zur Dauer von $V_1$ ausgedrückt wird. Der Grund dafür liegt in einer Synchronisierung des Gipfels mit dem Offset von $V_1$, ähnlich wie in der Untersuchung von Lippus et al. (2013). Dennoch zeigen sich in der vorliegenden Studie ähnliche Gipfelsynchronisierungsunterschiede wie bei VQ-Wörtern, wenn der Gipfel in Proportion zur $V_1K_2$-Dauer ausgedrückt wird. Sowohl in VQ- als auch KQ-Wörtern hat die Gipfelsynchronisierung proportional zur $V_1K_2$-Dauer die Reihenfolge Q3 $<$ Q2 $<$ Q1, d.h. der Gipfel von Q3-Wörtern ist früher synchronisiert als der Gipfel von Q2-Wörtern usw.

Die Ergebnisse der aktuellen Studie legen die Vermutung nahe, dass der Gipfel von VQ- und KQ-Wörtern im Estnischen mit einem phonologischen Ankerpunkt aligniert sein könnte: dem Offset des ersten Mora des Worts (vgl. Abbildung 2.10). In VQ1-, KQ1-, KQ2- und KQ3-Wörtern ist der Gipfel mit dem $V_1$-Offset synchronisiert, das dem Offset des ersten Mora entspricht. In VQ2- und VQ3-Wörtern befindet sich das Mora-Offset irgendwo in $V_1$ und seine genaue Position kann nur durch einen Vergleich mit der proportionalen Gipfelsynchronisierung der KQ2- und KQ3-Wörter geschätzt werden. Da die proportionale Gipfelsynchronisierung von VQ2- und VQ3- vs. KQ2- und KQ3-Wörtern ähnlich ist, könnte man annehmen, dass auch in VQ2- und VQ3-Wörtern der Gipfel mit dem Offset des ersten Mora synchronisiert ist. Da jedoch das Mora-Konzept nicht unproblematisch ist für das Estnische, ist unter Umständen eine einfachere Interpretation der Ergebnisse zu bevorzugen: In VQ1- und KQ1-Wörtern ist der Gipfel mit dem Ende von $V_1$ synchronisiert, in VQ2- und KQ2-Wörtern mit einem Punkt, der kurz vor dem zeitlichen Mittelpunkt der $V_1K_2$-Sequenz liegt, und in VQ3- und KQ3-Wörtern mit einem Punkt im ersten Drittel der $V_1K_2$-Sequenz.

### A.1.2 Variation der Sprechgeschwindigkeit

Im dritten Experiment (Kapitel 4) dieser Dissertation wird untersucht, ob Sprechgeschwindigkeit einen Einfluss auf die Gipfelsynchronisierung hat. Dazu wurden die Zielwörter in Trägersätze eingebettet, in denen eine unbetonte Silbe auf das Zielwort folgt, um eine Gipfelverschiebung durch die Satzgrenze zu vermeiden. Die Versuchspersonen lasen die Sätze einmal in normalem und einmal in schnellem Sprechtempo. Es wurden insgesamt zwei Hypothesen getestet:

(1) Der Gipfel von VQ- und KQ-Wörtern ist ähnlich.

(2) Sprechgeschwindigkeit hat keinen Einfluss auf die Gipfelsynchronisierung proportional zur $V_1K_2$-Dauer, da die Gipfelsynchronisierung im Estnischen für die Unterscheidung zwischen den Quantitätsstufen verwendet wird.

Die Ergebnisse des aktuellen Experiments bestätigen die erste Hypothese. Die Gipfelsynchronisierung von VQ- und KQ-Wörtern ist ähnelt sich. Sowohl die absolute als auch die proportionale Gipfelsynchronisierung ist ähnlich wie im ersten Experiment. Wieder hat die proportionale Gipfelsynchronisierung die Reihenfolge Q3 < Q2 < Q1, d.h. der Gipfel von Q3-Wörtern ist früher synchronisiert als der von Q2-Wörtern usw. Ähnlich wie im ersten Experiment unterscheidet sich die Gipfelsynchronisierung von VQ2-Wörtern von den Ergebnissen früherer Studien zum Estnischen, in denen der Gipfel von VQ2-Wörtern gegen Ende von $V_1$ gefunden wurde (Asu et al., 2009; Krull, 1992, 1993; Lehiste, 1960, 1997a,b, 2003; Liiv, 1961; Lippus et al., 2013). Ähnlich wie in Krull (1997) befindet sich der Gipfel von VQ2-Wörtern auch im aktuellen Experiment eher in der Mitte von $V_1$ und nicht am Ende. Der Grund dafür könnte in der Verwendung des Wendepunkts (*turning point*) in den Studien von Asu et al. (2009) und Lippus et al. (2013) liegen, der später aligniert ist als der Gipfel, der in der aktuellen Studie und in Krull (1997) verwendet wird.

Auch die zweite Hypothese kann bestätigt werden: Sprechgeschwindigkeit hat keinen Einfluss auf die Gipfelsynchronisierung proportional zur $V_1K_2$-Dauer, ähnlich wie in Studien zum Russischen (Igarashi, 2004), Englischen (Ladd et al., 1999), Mandarinchinesischem (Xu, 1998) und Thai (Nitisaroj, 2006).

## A.2 Phrasenfinale Längung im Estnischen

Die Ergebnisse des ersten Experiments (Kapitel 2) zeigen, dass die Segmentdauern der zweisilbigen Zielwörter unterschiedlich von der herannahenden Satzgrenze beeinflusst werden. Es wird vermutet, dass phrasenfinale Längung (PFL) im Estnischen einen unterschiedlichen Einfluss auf VQ- vs. KQ-Wörter hat. Ziel des zweiten Experiments (Kapitel 3) ist es, dies näher zu untersuchen.

Frühere Studien zu PFL im Estnischen konzentrieren sich hauptsächlich auf die Vokale von zweisilbigen VQ-Wörtern (Asu et al., 2009; Eek und Meister, 2003; Krull, 1997). Während eine Längung von $V_2$ stabil zu sein scheint, variiert die Längung von $V_1$ je nach Quantitätsstufe und Studie. Eek und Meister (2003) untersuchen zusätzlich auch die Konsonanten von VQ-Wörtern und finden eine Längung von $K_2$ in allen Quantitätsstufen.

Zur Erklärung von PFL in anderen Sprachen wurden bisher drei verschiedene Modelle verwendet: ein strukturbasiertes (*Structure-based*), ein inhaltsbasiertes (*Content-based*) und ein Hybrid (*Hybrid-view*) Modell (Turk und Shattuck-Hufnagel, 2007). Im strukturbasierten Modell längt PFL eine linguistische Struktur, z.B. den Reim der wortfinalen und der primär betonten Silbe wie im Amerikanischen Englischen (Turk und Shattuck-Hufnagel, 2007). Im inhaltsbasierten Modell wird der Teil der segmentellen Kette gelängt, der mit einer abstrakten prosodischen Geste überlappt, wie zum Beispiel der $\pi$-Geste ($\pi$-*gesture*) des $\pi$-Gesten-Konzepts (z.B. Byrd et al., 2006). Im Hybrid Modell wird angenommen, das PFL eine strukturell fixe Domäne der segmentellen Kette beeinflusst, z.B. den Reim der wortfinalen Silbe. Kann der Reim z.B. wegen phonologischer Quantitätskontraste nicht unbegrenzt gelängt werden, können zusätzlich auch frühere Segmente gelängt werden (Cambier-Langeveld, 1997).

Im Englischen (Turk und Shattuck-Hufnagel, 2007), im Finnischen (Nakai et al., 2009) und in einer Studie zu VQ-Wörtern des Estnischen (Eek und Meister, 2003) wurde progressive Längung (*progressive lengthening*) gefunden, d.h. je näher sich ein Segment zur hinteren Wortgrenze befindet, desto mehr wird es von PFL beeinflusst.

In Bezug auf PFL werden im aktuellen Experiment zwei Hypothesen getestet:

(1) Der Reim der ersten und der zweiten Silbe sind die Domänen von PFL im Estnischen und können am besten mit einem strukturbasierten Modell beschrieben werden.

(2) Im Estnischen tritt sowohl bei VQ- als auch KQ-Wörtern progressive Längung auf.

Für die Analyse wurden dieselben Daten wie im ersten Experiment (Kapitel 2) verwendet, d.h. auf die Zielwörter folgten entweder zwei (nicht-finaler Kontext) oder keine (finaler Kontext) unbetonten Silben.

Das aktuelle Experiment zeigt, dass PFL im Estnischen mit den Quantitätsstufen und -typen interagiert. Während in VQ1- und KQ1-Wörtern PFL $V_1$ und $K_2$ signifikant beeinflusst, werden in VQ2- und VQ3-Wörtern $V_1$ und $V_2$ gelängt und in KQ2- und KQ3-Wörtern $K_2$ und $V_2$.

Die Hypothese, dass die Domäne von PFL im Estnischen der Reim der ersten und zweiten Silbe ist, kann nur für VQ2- und VQ3-Wörter bestätigt werden. Dennoch ist ein strukturbasiertes Modell für die Erklärung von PFL im Estnischen am besten geeignet. Allerdings sind die linguistischen Strukturen, die von PFL beeinflusst werden, nicht wie im Amerikanischen Englischen (Turk und Shattuck-Hufnagel, 2007) die Silbenreime, sondern der Hauptträger des Quantitätskontrasts, d.h. $V_1$ in VQ-Wörtern und $K_2$ in KQ-Wörtern. Abhängig von Quantitätsstufe und -typ werden zusätzlich noch weitere Segmente gelängt.

Auch die zweite Hypothese konnte bestätigt werden. Ähnlich wie im Amerikanischen Englischen (Turk und Shattuck-Hufnagel, 2007), in einer früheren Studie zu estnischen VQ-Wörtern (Eek und Meister, 2003) und im Finnischen (Nakai et al., 2009) findet sich im aktuellen Experiment progressive Längung, d.h. je näher ein Segment an der Satzgrenze ist, desto mehr wird es gelängt. Progressive Längung betrifft im aktuellen Experiment nur VQ2-, VQ3- und KQ1-Wörter und es wird angenommen, dass pogressive Längung nur dann auftritt, wenn die gelängten Segmente nicht wortfinal sind und nicht unmittelbar aufeinander folgen.

## A.3 Allgemeine Diskussion

Nach z.B. Lehiste (2003) sind die Unterschiede in der Gipfelsynchronisierung zwischen VQ2- und VQ3-Wörtern in einem Lautwandel zu finden, der stattfand als sich das Estnische vom Finnischen trennte. Trotzdem sind die Gründe für die Gipfelsynchronisierungsunterschiede immer noch unklar. Die Gipfelsynchronisierungsunterschiede, die in dieser Dissertation gefunden wurden, lassen sich mit Hilfe der segmentellen Verankerungshypothese (*segmental anchoring hypothesis*) erklären, die besagt, dass tonale Ziele mit bestimmten Ankerpunkten in der segmentellen Kette synchronisiert sind (z.B. Ladd et al., 1999, 2000; Schepman et al., 2006). Als Ankerpunkt für das Estnische könnte das Offset des ersten Mora dienen. Unterschiede in der poportionalen Gipfelsynchronisierung kommen zustande, indem der Ankerpunkt mit den temporalen Korrelaten des Quantitätskontrasts interagiert.

Sowohl im ersten als auch im dritten Experiment dieser Disseration wird der Einfluss von Zeitdruck (*time pressure*) auf die Gipfelsynchronisierung untersucht, d.h. durch prosodische Faktoren bleibt weniger Zeit den Intonationsgipfel zu synchronisieren. Im ersten Experiment entsteht der Zeitdruck, da die Anzahl an unbetonten Silben nach dem Zielwort abnimmt. Im zweiten Experiment führt eine schnellere Sprechgeschwindigkeit zu einem Zeitdruck. Nur die erste Bedingung beeinflusste die Gipfelsynchronisierung. Der Grund könnte das sein, was Ladd (2008) einen phonologische vs. phonetischen Grund für den Zeitdruck nennt. Nach Ladd (2008) wird die Gipfelsynchronisierung nur von einem Zeitdruck mit einem phonologischen Grund (z.B. die Anzahl an Silben zwischen einem Tonakzent und einer prosodischen Grenze) beeinflusst, nicht aber von einem Zeitdruck phonetischer Natur (z.B. die tatsächliche Dauer zwischen einem Tonakzent und einer prosodischen Grenze). Die Ergebnisse der vorliegenden Dissertation bestätigen die Vermutung von Ladd (2008). Ein phonologischer Zeitdruck aufgrund einer Variation der Anzahl an unbetonten Silben hat einen Einfluss auf die Gipfelsynchronisierung, wohingegen Sprechgeschwindigkeit als phonetischer Zeitdruck die Gipfelsynchronisierung nicht beeinflusst.

In dieser Dissertation wird gezeigt, dass die Satzgrenze im Estnischen nicht nur die Gipfelsynchronisierung beeinflusst, sondern auch die Segmentdauern. Der Einfluss der phrasenfinalen Längung (PFL) hängt sehr von Quantitätsstufe und -typ ab. Dennoch gibt es eine linguistische Struktur, die in allen Fällen von PFL beeinflusst wird: der Hauptträger des Quantitätskontrasts, d.h. $V_1$ in VQ-Wörtern und $K_2$ in KQ-Wörtern. Dieses Ergebnis kann im Sinne eines strukturbasierten Modells zur Erklärung von PFL interpretiert werden. Neben dem Hauptträger der Quantität sind allerdings auch noch andere Segmente von PFL betroffen. Deshalb kann ein strukturbasiertes Modell, das nur eine linguistische Struktur beinhaltet, für das Estnische nicht bestätigt werden. Trotzdem ist ein strukturbasiertes Modell das geeignetste Modell für das Estnische.

# Bibliography

Arvaniti, A. and Baltazani, M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In Jun, S.-A., editor, *Prosodic Typology: The Phonology of Intonation and Phrasing*, pages 84–117. Oxford University Press, Oxford.

Arvaniti, A. and Ladd, D. R. (1995). Tonal alignment and the representation of accentual targets. In *Proceedings of the XIIIth International Congress of Phonetic Sciences ICPhS'95, Stockholm*, pages 220–223.

Arvaniti, A., Ladd, D. R., and Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, 26:3–25.

Asu, E. and Nolan, F. (2007). The analysis of low accentuation in Estonian . *Language and speech*, 50(4):567–588.

Asu, E. L. (2004). *The phonetics and phonology of Estonian intonation.* PhD thesis, University of Camebridge.

Asu, E. L. (2005). Towards a model of Estonian intonation. In Langemets, M. and Penjam, P., editors, *The second baltic conference on human language technologies. Proceedings, April 4-5, 2005, Tallinn, Estonia*, pages 95–100.

Asu, E. L. (2006). Rising intonation in Estonian: an analysis of map task dialogues and spontaneous conversations. In *Fonetiikan Päivät 2006 / The phonetics symposium 2006, Helsinki University*, pages 1–8.

Asu, E. L., Lippus, P., and Teras, P.; Tuisk, T. (2009). The realization of Estonian quantity characteristics in spontaneous speech. In Vainio, M., Aulanko, R., and Aaltonen, O., editors, *Nordic Prosody - Proceedings of the Xth conference, Helsinki 2008*, pages 49–56. Peter Lang Frankfurt.

Asu, E. L. and Nolan, F. (1999). The effect of intonation on pitch cues to the Estonian quantity contrast. In *Proceedings of the 14th International Congress of Phonetic Sciences. San Francisco, 3*, pages 1873–76.

Asu, E. L. and Nolan, F. (2006). Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. In *Speech Prosody 2006. Dresden*, pages 249–252.

Asu, E. L. and Teras, P. (2009). Estonian. *Journal of the International Phonetic Association*, 39:367–372.

Atterer, M. and Ladd, R. D. (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics*, 32.

Beckman, M. and Ayers Elam, G. (1997). Guidelines for ToBI Labelling (version 3.0, March 1997). Obtain by writing to tobi@ling.ohio-state.edu.

Beckman, M. and Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3:255–309.

Boersma, P. and Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.2.14, retrieved 10 July 2010 from http://www.praat.org/.

Bruce (1977). *Swedish word accents in sentence perspective*. Gleerup, Lund.

Bye, P. (1997). A generative perspective on overlength in Estonian and Saami. In Lehiste, I.; Ross, J., editor, *Estonian Prosody: Papers from a Symposium*, pages 36–68. Tallinn: Institute of Estonian Language.

Byrd, D., Krivokapić, and Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America*, 120(3):1589–1599.

Byrd, D. and Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lenghtening. *Journal of Phonetics*, 31(2):149–180.

Cambier-Langeveld, T. (1997). The domain of final lengthening in the production of Dutch. In de Hoop, H. and Coerts, J., editors, *Linguistics in the Netherlands*, pages 13–24. John Benjamins, Amsterdam.

Caspers, J. and van Heuven, V. J. (1993). Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica*, 50:161–171.

Collinder, B. (1929). *Über den finnisch-lappischen Quantitätswechsel*. Uppsala.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge University Press, London.

Dalton, M. and Ní Chasaide, A. (2005). Tonal alignment in Irish Dialects. *Language and speech*, 48:441–464.

De la Mota, C. (2005). Alignment, word boundaries and speech rate in Castilian Spanish. In *Paper presented the Second Conference on Phonetics and Phonology in Iberia, PaPI 2005, Bellaterra, June 20-21*.

Dilley, L. and Brown, M. (2007). Effects of pitch range variation on F0 extrema in an imitation task. *Journal of Phonetics*, 35:523–551.

Dilley, L., Ladd, D. R., and Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics*, 33:115–119.

D'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment*. PhD thesis, Department of Linguistics. The Ohio State University.

D'Imperio, M., Petrone, C., and Nguyen, N. (2007). Effects of tonal alignment on lexical identification in Italian. In Gussenhoven, C. and Riad, T., editors, *Tones and Tunes*. Mouton de Gruyter, Berlin.

D'Imperio, P. (2006). Preface - Current Issues in Tonal Alignment. *Rivista di Linguistica*, 18.1:3–18.

Draxler, C. and Jänsch, K. (2008). Wikispeech – a content management system for speech databases. In *Proc. Interspeech*, pages 1646–1649, Brisbane.

Draxler, C. and Jänsch, K. (2004). SpeechRecorder – a Universal Platform Independent Multi-Channel Audio Recording Software. In *Proc. of LREC*, pages 559–562, Lisbon.

Eek, A. (1974). Observations on the duration of some word structures I. *Estonian Papers in Phonetics*, pages 31–57.

Eek, A. and Help, T. (1987). The interrelationship between phonological and phonetic sound change: a great rhythm shift of Old Estonian. In *Proceedings of the 11th International Congress of Phonetics Sciences. Tallinn.*

Eek, A. and Meister, E. (1997). Simple perception experiments on Estonian word prosody: foot structure vs. segmental quantity. In Lehiste, I.; Ross, J., editor, *Estonian Prosody: Papers from a Symposium*, pages 71–99. Tallinn: Institute of Estonian Language.

Eek, A. and 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. and Meister, E. (2003). Foneetilisi katseid ja arutlusi kvantiteedi alalt (I). *Keel ja Kirjandus*, 11:815–837.

Ehala, M. (2003). Estonian Quantity: Implications for Moraic Theory. In Nelson, D. and Manninen, S., editors, *Generative Approaches to Finnic and Saami Linguistics*. Stanford.

Gili Fivela, B. (2002). Tonal alignment in two Pisa Italian peak accents. In *Proceedings of Speech Prosody 2002, Aix-en-Provence*, pages 339–342.

Gili Fivela, B. and Savino, M. (2003). Segments, Syllables and Tonal Alignment: a Study on Two Varieties of Italian. In *Proceedings of the XVth International Congress of Phonetic Sciences, Barcelona, Spain*, pages 2933–2936.

Grabe, E. and Low, E. L. (2002). Durational Variability in Speech and the Rhythm Class hypothesis. *Laboratory Phonology*, 7:515–546.

Grice, M. and Baumann, S. (2007). An Introduction to Intonation - Functions and Models. In Trouvain, J. and Gut, U., editors, *Non-Native Prosody. Phonetic Description and Teaching Practice*. De Gruyter (=Trends in Linguistics. Studies and Monographs [TiLSM] 186), Berlin, New York.

Grice, M., Baumann, S., and Benzmüller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In *Prosodic Typology. The Phonology of Intonation and Phrasing*, pages 55–83. Oxford University Press, Oxford.

Hakokari, J., Saarni, T., Salakoski, T., Isoaho, J., and Aaltonen, O. (2005). Determining prepausal lengthening for Finnish rule-based speech synthesis. In *Proceedings of Speech Analysis, Synthesis and Recognition: Applications of Phonetic (SASR 2005), Kraków.*

Halliday, M. A. K. (1967). *Intonation and Grammar in British English.* Mouton, Paris.

Harrington, J. (2010). *The Phonetic Analysis of Speech Corpora.* Wiley-Blackwell.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20(2):253–306.

Hayes, B. (1995a). *Metrical Stress Theory: Principals and case studies.* The University of Chicago Press.

Hayes, B. (1995b). Weight of CVC can be determined by context. In Cole, J. and Kisseberth, C., editors, *Perspectives in Phonology*, pages 61–79. Stanford, CA.

Hellmuth, S. (2005). Pitch accent alignment in Egyptian Arabic: Exploring boundaries of cross-linguistic alignment variation. In *Paper presented at II phonetics and phonology in Iberia, PaPi 2005, Bellaterra, June 20-21.*

Hellmuth, S. (2006). *Intonational pitch accent distribution in Egyptian Arabic.* PhD thesis, Ph.D. School of Oriental and African Studies, University of London.

Hint, M. (1980). Minevikuline ja tulevikuline aines keelesüsteemis. Prosoodiatüübi nihked ja selle tagajärjed. *Keel ja Kirjandus*, 24:215–223, 270–278, 349–355.

Igarashi, Y. (2004). 'Segmental anchoring' of F0 under changes in speech rate: evidence from Russian. In *Speech Prosody 2004*, pages 25–28.

Ishihara, T. (2003). A phonological effect on tonal alignment in Tokyo Japanese. In *Proceedings of the XVth international congress of phonetic sciences*, volume 1, pages 615–618.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) Labelling Conventions. *UCLA Working Papers in Phonetics*, 99:149–173.

Kager, R. (1994). Ternary rhythm in alignment theory. Ms., University of Utrecht. ROA-35.

Keevallik, L. (2003). Terminally Rising Pitch Contours of Response Tokens in Estonian . *Crossroads of Language, Interaction, and Culture*, 5:49–65.

Kingdon, R. (1958). *The groundwork of English intonation.* Longman, London.

Kleber, F. and Rathcke, T. (2008). More on the "segmental anchoring" of prenuclear rises. In *Speech Prosody 2008, Campinas, Brazil*, pages 583–586.

Kohler, K. (1987). Categorical pitch perception. In *Proceedings of the 11th International Congress of Phonetics Sciences. Tallinn.*

Kohler, K. (1991). Terminal Intonation Patterns in Single-Accent Utterances of German: Phonetics, Phonology and Semantics. *AIPUK*, 25:115–185.

Kohler, K. (2005). Timing and communicative functions of pitch contours. *Phonetica*, 62:88–105.

Krull, D. (1991). Stability in some Estonian duration relations. *PERILUS 13, Institute of Linguistics, University of Stockholm*, pages 57–60.

Krull, D. (1992). Temporal and tonal correlates to quantity in Estonian. *PERILUS 15, Institute of Linguistics, University of Stockholm*, pages 17–36.

Krull, D. (1993). Word-prosodic features in Estonian conversational speech: some preliminary results. *Prosody-1993*, pages 148–151.

Krull, D. (1997). Prepausal Lengthening in Estonian: Evidence from Conversational Speech. In Lehiste, I.; Ross, J., editor, *Estonian Prosody: Papers from a Symposium*, pages 136–148. Tallinn: Institute of Estonian Language.

Krull, D. (1999). Foot isochrony in Estonian. In *Proceedings of the 14th International Congress of Phonetic Sciences. San Francisco, 3*, pages 1063–1066.

Krull, D. (2001). Perception of Estonian word prosody in whispered speech. In Dommelen van, W. A. and Fretheim, T., editors, *Nordic Prosody: Proceedings of the VIIIth Conference, Trondheim, August 19-21, 2000*, pages 153–164.

Kubozono, H. (1996). Speech segmentation and phonological structure. In Otake, T. and Cutler, A., editors, *Phonological structure and language processing: cross-linguistic studies*, pages 77–94. Mouton de Gruyter.

Laanekask, H. and Erelt, T. (2003). Written Estonian. In Erelt, M., editor, *The Estonian language. Linguistica Uralica. Supplementary Series I*, pages 273–342. Eesti Teaduste Akadeemia Kirjastus, Tallinn.

Ladd, D., Schepman, A., White, L., and May Quarmby, L. (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics*, 37:145–161.

Ladd, D. R. (1996). *Intonational Phonology*. Camebridge University Press.

Ladd, D. R. (2008). *Intonational Phonology*. Camebridge University Press, Camebridge, 2nd edition.

Ladd, D. R., Mennen, I., and Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes of speech rate. *Journal of the Acoustic Society of America 80*, pages 1543–1554.

Ladd, D. R., Mennen, I., and Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustic Society of America 80*, pages 2685–2696.

Ladd, R. D. (1983). Phonological features of intonational peaks. *Language*, pages 721–759.

Lecumberri, M. L. G. (1997). Traditional British analyses of intonation: a review. *Atlantis*, 19(2):103–112.

Lehiste, I. (1960). Segmental and syllabic quantity in Estonian. In Sebeok, T. A., editor, *American Studies in Uralic Linguistics. Uralic and Altaic Seris 1*, pages 21–82. Bloomington.

Lehiste, I. (1966). *Consonant Quantity and Phonological Units in Estonian*. Bloomington, India University.

Lehiste, I. (1970-1975). Experiments with synthetic speech concerning quantity in Estonian. In *Congressus Tertius internationalis Fenno-Ugristarum, Tallinae habitus 17-23. VIII 1970. Pars I: Acta Linguistica*, pages 254–269.

Lehiste, I. (1981). Sentence and paragraph boundaries in estonian. In *Congressus Quintus Internationalis Fenno-Ugristarum, Turku, 20.-27. VIII. 1980. Pars VI*, pages 164–169.

Lehiste, I. (1990). Moraic analysis of Finnish and Estonian. In Jakab, L., Keresztes, L., Kiss, A., and Maticsak, S., editors, *Congressus Septimus Internationalis Fenno-Ugristarum, Debrecen 27.VIII – 2.IX. 1990. Sessiones Sectionum 3b, Linguistics*, pages 286–291.

Lehiste, I. (1997a). Search for phonetic correlates in Estonian prosody. In Lehiste, I.; Ross, J., editor, *Estonian Prosody: Papers from a Symposium*, pages 13–35. Tallinn: Institute of Estonian Language.

Lehiste, I. (1997b). The structure of trisyllabic words. In Lehiste, I.; Ross, J., editor, *Estonian Prosody: Papers from a Symposium*, pages 149–163. Tallinn: Institute of Estonian Language.

Lehiste, I. (2003). Prosodic change in progress: from quantity language to accent language. In Fikkert, P. and Jacobs, H., editors, *Development in Prosodic Systems*, pages 47–65. Berlin, New York: Mouton de Gruyter.

Lehiste, I. and Fox, R. A. (1992). Perception of prominence by Estonian and English listeners. *Language and Speech*, 35:419–434.

Liberman, M. (1975). *The intonational system of English.* PhD thesis, MIT, Massachusets.

Liiv, G. (1961). Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid. *Keel ja Kirjandus*, 7-8:412–424 and 480–490.

Lindström, L. (2006). Infostruktuuri osast eesti keele sõnajärje muutumisel. *Keel ja Kirjandus*, 11:875–888.

Lippus, P. (2010). Variation in Vowel Quality as a Feature of Estonian Quantity. In Hasegawa-Johnson, M., Bradlow, A., Cole, J., Livescu, K., Pierrehumbert, K., and Shin, C., editors, *Speech Prosody 2010. Chicago, USA*.

Lippus, P., Asu, E. L., Teras, P., and Tuisk, T. (2013). Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics*, 41:17–28.

Lippus, P. and Pajusalu, K. (2009). Regional variation in the perception of Estonian quantity. In Vainio, M., Aulanko, R., and Aaltonen, O., editors, *Nordic Prosody - Proceedings of the Xth Conference, Helsinki 2008*, pages 151–157.

Lippus, P., Pajusalu, K., and Allik, J. (2007). The Tonal Component in Perception of the Estonian Quantity. In *The Proceedings of the 16th International Congress of Phonetic Sciences: 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 6-10 August 2007*, pages 1049–1052.

Lippus, P., Pajusalu, K., and Allik, J. (2009). The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics*, 37:388–396.

Lippus, P., Pajusalu, K., and Allik, J. (2011). The role of pitch cue in the perception of the Estonian long quantity. In Frota, S., Elordieta, G., and Prieto, P., editors, *Prosodic categories: Production, Perception and Comprehension*, pages 231–242. Springer.

McCarthy, J. and Prince, A. (1986). Prosodic Morphology. unpublished manuscript, University of Massachusetts.

Meister, E. and Meister, L. (2011). Perception of the short vs. long phonological category in Estonian by native and non-native listeners. *Journal of Phonetics*, 39(2):212–224.

Meister, E., Werner, S., and Meister, L. (2011). Short vs. long category perception affected by vowel quality. In *Proceedings of the ICPhS XVII 17-21 August 2011 Hong Kong*, pages 1362–1365.

Mihkla, M. (2005). Modelling pauses and boundary lengthening in synthetic speech. In *Proceedings of the Second Baltic Conference on Human Language Technologies, April 4-5 2995, Tallin*, pages 305–310.

Mihkla, M. (2006). Pausid kõnes. *Keel ja Kirjandus*, XLIX(4):286–295.

Möbius, B. and Jilka, M. (2007). Effects of syllable structure and nuclear pitch accents on peak alignment: a corpus-based analysis. In *The Proceedings of the 16th International Congress of Phonetic Sciences: 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 6-10 August 2007*, pages 1173–1176.

Mücke, D., Grice, M., Becker, J., and Hermes, A. (2009). Sources of variation in tonal alignment: Evidence from acoustic and kinematic data. *Journal of Phonetics*, 37:321–338.

Mücke, D. and Hermes, A. (2007). Phrase boundaries and peak alignment: an acoustic and articulatory study. In *The Proceedings of the 16th International Congress of Phonetic Sciences: 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 6-10 August 2007*, pages 997–1000.

Nakai, S., Kunnari, S., Turk, A., Suomi, K., and Ylitalo, R. (2009). Utterance final lengthening and quantity in Northern Finnish. *Journal of Phonetics*, 37:29–45.

Niebuhr, O. (2007). The signalling of German rising-falling intonation categories - the interplay of synchronization, shape, and height. *Phonetica*, 64:174–193.

Nitisaroj, R. (2006). Thai tonal contrast under changes in speech rate and stress. In *Speech Prosody 2006, Dresden, Germany, May 2-5 2006*.

Nolan, F. (2006). Intonation. In Aarts, B. and McMahon, A. M. S., editors, *The handbook of English linguistics*. Blackwell, Oxford.

Nolan, F. and Asu, E. L. (2009). The Pairwise Variability Index and Coexisting Rhythms in Language. *Phonetica*, 66(1-2):64–77.

Pajusalu, K. (2003). Estonian dialects. In Erelt, M., editor, *The Estonian language. Linguistica Uralica. Supplementary Series I*, pages 231–272. Eesti Teaduste Akadeemia Kirjastus, Tallinn.

Peters, J. (1999). The timing of nuclear high accents in German dialects. In *Proceedings of the 14th International Congress of Phonetic Sciences. San Francisco, 3*, pages 1877–80.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT.

Pierrehumbert, J. and Beckman, M. (1988). *Japanese Tone Structure*. The MIT Press.

Pierrehumbert, J. and Steele, S. (1989). Categories of Tonal Alignment in English. *Phonetica*, 46:181–196.

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

Plüschke, M. (2009). The influence of sentence stress on disyllabic words containing short, long and overlong consonants in Estonian . Unpublished Magister's thesis. Ludwig-Maximilians-Universität Munich.

Plüschke, M. (2011). Peak alignment in falling accents in Estonian. In *Proceedings of the ICPhS XVII 17-21 August 2011 Hong Kong*, pages 1614–1617.

Prieto, P. (2005). Stability effects in tonal clash contexts in Catalan. *Journal of Phonetics*, 33:215–242.

Prieto, P. (2009). Tonal alignment patterns in Catalan nuclear falls. *Lingua*, special issue on The phonology and phonetics of tone from a typological perspective, ed. by F. Kügler, L. Downing, and S. Zerbian(119):865–880.

Prieto, P. (in press). The Intonational Phonology of Catalan. In Jun, S.-A., editor, *Prosodic Typology 2*. Oxford University Press, Oxford.

Prieto, P. and Torreira, F. (2007). The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics*, 35:473–500.

Prieto, P., van Santen, J., and Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23:429–451.

Prince, A. (1980). A metrical theory for Estonian quantity. *Linguistic Inquiry*, 11(3):511–562.

Rathcke, T. and Harrington, J. (2006). Is there a distinction between H+!H* and H+L* in standard German?Evidence from an acoustic and auditory analysis. In *Proceedings of the International Speech Prosody Conference, Dresden.*

Rathcke, T. and Harrington, J. (2007). The phonetics and phonology of high and low tones in two falling f0-contours in Standard German. In *Interspeech 2007, Antwerpen*, pages 982–985.

Ross, J. and Lehiste, I. (2001). *The temporal structure of Estonian Runic Songs*. Mouton de Gruyter, Berlin, New York.

Sahkai, H., Kalvik, M.-L., and Mihkla, M. (2013). Prosodic effects of Information Structure in Estonian. In Asu, E. L. and Lippus, P., editors, *Nordic Prosody - Proceedings of the XIth conference, Tartu 2012*, Frankfurt am Main. Peter Lang.

Salveste, N. (2011). Variation of pitch cues in the perception of Estonian quantities. In *XXVI Fonetiikan päivät 2010*, pages 42–46.

Salveste, N. (2013). Focus perception in Estonian: Is it governed by syntax or by prosody? In Asu, E. L. and Lippus, P., editors, *Nordic Prosody - Proceedings of the XIth conference, Tartu 2012*, Frankfurt am Main. Peter Lang.

Schaefer-Vincent, K. (1983). Pitch period detection and chaining: Method and evaluation. *Phonetica*, 40:177–202.

Schepman, A., Lickley, R., and Ladd, R. D. (2006). Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics*, 34:1–28.

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech. In *Proceedings of the 14th International Congress of Phonetic Sciences. San Francisco, 3*, pages 607–610.

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992). ToBI: a standard for labeling English prosody. In *Proceedings of the Second International Conference on Spoken Language Processing, Banff, Canada*, pages 867–870.

Silverman, K. and Pierrehumbert, J. (1990). The Timing of Prenuclear High Accents in English. *Papers in Laboratory Phonology I*, pages 72–106.

Steele, S. A. (1986). Nuclear accent F0 peak location: Effects of rate, vowel, and number of following syllables. *Journal of the Acoustic Society of America 80 (Supplement 1)*, page S51.

't Hart, J., Collier, R., and Cohen, A. (1990). *A perceptual stduy of intonation. An experimental-phonetic approach of speech melody*. Cambridge University Press, Camebridge.

Tael, K. (1988). *Sõnajärjemallid eesti keeles (võrrelduna soome keelega). Preprint KKI-56*. Eesti NSV Teaduste Akadeemia. Keele ja Kirjanduse Instituut, Tallinn.

Trager, G. L. and Smith, H. L. (1951). *An outline of English structure*. Battenburg Press. Reprinted 1957 by American Council of Learned Societies, Washington, Normann, OK.

Turk, A. E. and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35:445–472.

Venditti, J. J. (2005). The JToBI model of Japanese intonation. In Jun, S.-A., editor, *Prosodic Typology and Transcription: A Unified Approach*, pages 172–200. Oxford University Press, Oxford.

Veske, M. (1879). *Eesti keele healte õpetus ja kirjutuse viis*. Tartu.

Viitso, T.-R. (1998). Estonian. In Abondolo, D., editor, *The uralic languages*, pages 115–148. Routledge, London, New York.

Viitso, T.-R. (2003a). Rise and development of the Estonian language. In Erelt, M., editor, *The Estonian language. Linguistica Uralica. Supplementary Series I*, pages 130–162. Eesti Teaduste Akadeemia Kirjastus, Tallinn.

Viitso, T.-R. (2003b). Structure of the Estonian language. In Erelt, M., editor, *The Estonian language. Linguistica Uralica. Supplementary Series I*, pages 9–92. Eesti Teaduste Akadeemia Kirjastus, Tallinn.

Warner, N. and Arai, T. (2001). Japanese Mora-Timing: A review. *Phonetica*, 58:1–25.

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, 34:343–371.

Welby, P. and Lœvenbruck, H. (2005). Segmental "anchorage" and the French late rise. In *Proceedings of Interspeech 2005: the 9th Annual Conference on Speech Communication and Technology, Lisbon, Portugal*, page 2369–2373.

Welby, P. and Lœvenbruck, H. (2006). Anchored down in Anchorage: Syllable structure and segmental anchoring in French. *Rivista di Linguistica/ Italian Journal of Linguistics*, 18:74–124.

Wells, R. (1945). The pitch phonemes of English. *Language*, 21:27–40.

White, L. and Mády, K. (2008). The long and the short and the final: phonological vowel length and prosodic timing in Hungarian. In *Speech Prosody 2008, Campinas, Brazil*, pages 363–366.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. (1992). Segmental durations in the vicinity of prosodic phrases. *Journal of the Acoustical Society of America*, 91(3):1707–1717.

Xu, Y. (1998). Consistency of Tone-Syllable Alignment across Different Syllable Structures and Speaking Rates. *Phonetica*, 55:179–203.