The Estonian Swedish diphthongs /ai au oi ui/

Acoustic characteristics and cross-dialectal variation

Otto Ewald

Supervisor: Susanne Schötz
Centre for Language and Literature, Lund University
MA in Languages and Linguistics, Phonetics
SPVR01 Languages and Linguistics: Degree Project - Master's (Two Years) Thesis, 30 credits

June 2016

Abstract

This study investigates the acoustic characteristics and cross-dialectal variation of the diphthongs /ai au oi ui/ in the Nuckö, Ormsö, and Rickul varieties of Estonian Swedish, which have previously only been described impressionistically (Lagman, 1979; Danell, 1905-34). Eight acoustic measures, reflecting the F_1 and F_2 frequency values at the first (InF_1 and InF_2) and second ($FinF_1$ and $FinF_2$) vowel targets, total diphthong duration, amount of spectral change (TL), and spectral rate of change (TL_{roc} and VSL_{roc}), were employed to analyze differences between diphthong categories within each dialect as well as differences in the realization of each diphthong category across dialects.

The results, based on data from 498 diphthong tokens in 20 elderly speakers, indicate that the main realizations of /ai oi ui/ are cross-dialectally stable as [a͡ɪ]~[a͡e], [o͡ɪ]~[ɔ͡ɪ], and [u͡i] in all three dialects with major distinctions between diphthong categories found in the InF_1, InF_2, $FinF_1$, and TL_{roc} measures. In contrast, the results for /au/ indicate considerable cross-dialectal variation with the main realizations [a͡u̹]~[a͡ʉ], [œ͡ʏ], and [a͡u̹] in the Nuckö, Ormsö, and Rickul dialects respectively, which is reflected as variation in the height of the first target (InF_1), the backness of the second target ($FinF_2$), and the amount of spectral change (TL). Additionally, substantial intra-dialectal variation was also observed in the second target of Nuckö /au/ and in the first target of Ormsö /au/. Thus, for the first time, the current study provides acoustic evidence that confirms many of the observations featured in earlier impressionistic descriptions (Lagman, 1979; Danell, 1905-34) on the characteristics and cross-dialectal variation of the Estonian Swedish diphthongs /ai au oi ui/.

Keywords: Estonian Swedish, diphthongs, cross-dialectal variation, acoustic analysis

Acknowledgements

I would like to thank my supervisor Susanne Schötz for her support and feedback during my research and writing process, and for helping me focus on what is important and keeping me in the right direction throughout my work.

I would also like to express my appreciation for the help and advice offered by Joost van de Weijer on the statistical analysis that serves as the backbone of my investigation.

I would furthermore like to acknowledge everyone involved in the *Estlandssvenskans språkstruktur* (ESST) project for allowing me to base my study on data from the recordings they have collected, as well as for giving me the opportunity to partake in their field trips.

Additionally, I would also like to thank all informants, both at the Estonian Swedish community *Svenska Odlingens Vänner* (SOV) in Stockholm and elsewhere, without whose participation this study would not have been possible.

Finally, I would like to express my gratitude to my friends and family for their support and encouragement during my work on this thesis, especially in times of struggle.

Table of Contents

List of Figures

List of Tables

1 Introduction

Up until recently, the phonological system of Estonian Swedish has only been described impressionistically, notably in Danell (1905-1934) and Lagman (1979). However, several phonetic studies using modern acoustic methods have in the last couple of years targeted various parts of the Estonian Swedish phonological system, such as its long high vowels (Asu et al., 2009), word accents (Schötz & Asu, 2013; Schötz & Asu, 2015), laterals (Schötz et al., 2014; Asu et al., 2015a), and sibilants (Asu et al., 2015b).

The present study seeks to expand on this growing body of phonetic research on Estonian Swedish by investigating and describing the acoustic characteristics of the four diphthongs /ai au oi ui/ in the three Estonian Swedish dialects of Nuckö, Ormsö, and Rickul, which are commonly grouped together into one distinct dialect area (Lagman, 1979), using current acoustic methods for vowel analysis. More specifically, the study aims at targeting both the defining acoustic differences between each diphthong category within dialects, as well as any potential cross-dialectal variation in each of the diphthongs. Thus, this thesis constitutes the first acoustic study not only on Estonian Swedish diphthongs, but also on variation between different varieties within Estonian Swedish itself.

It should however be noted that the importance of describing Estonian Swedish does not only come from the lack of research on this variety, but also from the fact that the Estonian Swedish-speaking community is small and declining, at the present day consisting entirely of elderly speakers. It is therefore of utmost urgency to conduct linguistic studies on Estonian Swedish while the possibility to collect data to answer new research questions about this particular variety still exists. Additionally, the present study features an excellent opportunity to test recent acoustic methods, such as a method of formant extraction (Escudero et al., 2009) which have previously not been extensively used to collect acoustic data from diphthongs and spectro-temporal measures that can serve to characterize the dynamic aspects of diphthongs (Fox & Jacewicz, 2009).

The thesis is structured into 6 chapters. Chapter 2 gives a short overview of the different varieties of Estonian Swedish and their history as well as a background on diphthongs in Swedish, both in the standard language and in the dialects, with a particular focus on diphthongs in the Estonian Swedish dialects. This chapter also features an overview of some commonly used methodologies in the acoustic analysis of diphthongs. Chapter 3 restates and elaborates on the aims and research questions of the study, and furthermore describes the eight acoustic measurements employed in the analysis and the hypotheses that are connected to these measures based on earlier research.

Additionally, this chapter also gives an in-depth description of the data collection, preparation, and analysis procedures used in the current study. Chapter 4 subsequently presents an overview of the spectral characteristics of the diphthongs as well as the results of the analysis of each separate acoustic measure. In Chapter 5, the results for the different acoustic measures are then related to each other as well as to the research questions and hypotheses, and the findings are also discussed in the light of earlier research. Finally, Chapter 6 concludes the thesis by providing a summary of the study as a whole, and also gives suggestions for future research based on the current findings.

2 Background

2.1 Estonian Swedish

Estonian Swedish is an umbrella term for the collection of Swedish dialects that up until the present or near-present day have been spoken in the coastal areas of Estonia. These dialects have their origins in several immigration waves of Swedish-speaking people from both Sweden and Finland into Estonia over at least 300 years, starting in the early Middle Ages and eventually bringing many different varieties of Swedish into the area (Lagman, 1979).

Four main dialect areas, which are made up of the individual dialects shown in Figure 1, make up Estonian Swedish as a whole. In order from from the biggest to the smallest these dialect areas are Ormsö-Nuckö-Dagö (Est. Vormsi, Noarootsi, and Hiiumaa), Rågöarna-Vippal-Korkis (Est. Pakri saared, Vihterpalu, and Kurkse), Runö (Est. Ruhnu), and Nargö (Est. Naissaar) (Lagman, 1979). The interplay of several factors, such as historical isolation, varying origins, and a stable environment, has led to slow language change and the retention of distinct linguistic characteristics within each dialect area up until the present day (Lagman, 1979). The most pronounced differences between dialects are found mainly in their phonology and lexicon (Lagman, 1979). However, the dialects are not necessarily homogenous, and intra-dialectal variation occurs for example as local variations in the realization of some phonemic categories, such as diphthongs (Lagman, 1979; Danell, 1905-34).

In regards to contact with the neighboring Estonian language, such interactions have historically been the most extensive in Vippal, Korkis, and Nuckö, as well as in cities such as Hapsal (Est. Haapsalu) and Reval (present-day Tallinn). Even in the periphery of the Estonian Swedish-speaking area, Estonian Swedes abandoned their language in favor of Estonian rather than bringing Estonian influences into their language, minimizing the effect of language contact. However, in the last 20 years before the dissolution of the Estonian Swedish areas in conjunction with World War II, the contact with Estonian increased, partially due to it being introduced as an obligatory subject in schools (Lagman, 1979).

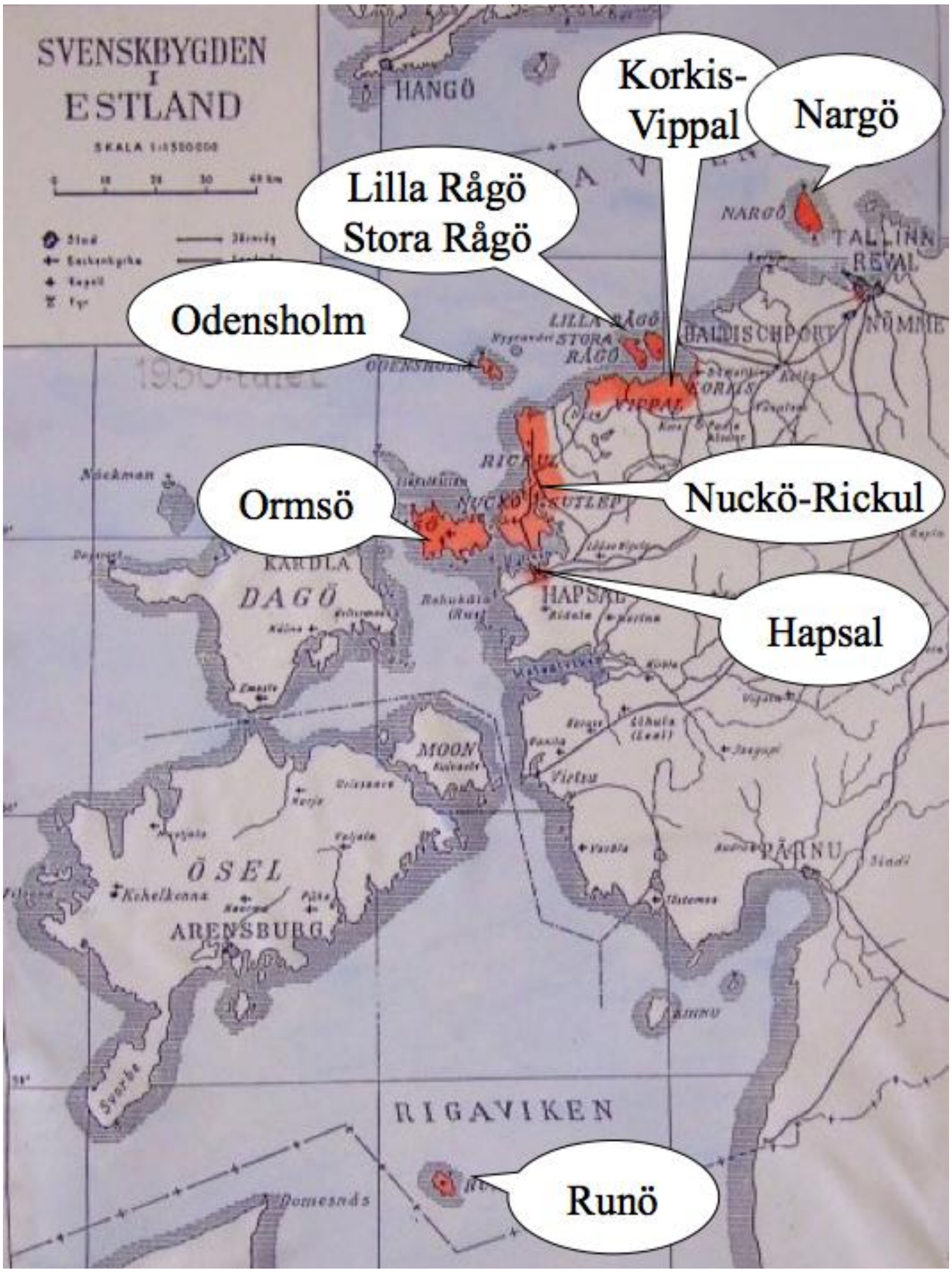

Figure 1. Geographical distribution of the Estonian Swedish dialects in the 1930s (adapted from Lagman, 1979). Red areas represent the approximate extent of the Estonian Swedish areas, and the labels represent the names of the areas where each dialect is spoken.

2.2 Diphthongs in Swedish

2.2.1 Diphthongs in Standard Swedish and Swedish dialects

Present-day Standard Swedish displays a large vowel inventory of 9 monophthongs that all come in pairs of long and short, although the merger of short /e/ and short /ɛ/ is common (Elert, 1989). While it can be argued that no phonological diphthongs exist in the language (Frid et al., 2012), the diphthong /au/ is considered by Bruce (2010) to exist in words such as *paus* /paus/ 'pause' and *nautisk* /nautisk/ 'nautical'. Furthermore, Riad (2014) describes two additional potential phonetic diphthongs that are represented as <eu> and <ou>~<oa> respectively in the orthography. The realization of orthographical <eu> varies between hiatus [eː.ʉ]~[eː.ɵ], diphthong [ɛ͡ʉ]~[ɛ͡ɵ], and vowel+consonant combination [ɛv] or [ɛf], and occurs in word such as *neutrum* [ˈneːʉ̣trɵm]~[ˈneːɵtrɵm]~[ˈnɛ͡ɵtrɵm] 'neuter' and *leukemi* [leˑʉ̣kẹˈmiː]~[lɛ͡ɵkẹˈmiː]~[lɛ̣fkẹˈmiː]~[lɛvˑkẹˈmiː] 'leukemia'. For orthographical <ou>~<oa>, the actual realization varies between a diphthong [ə͡u] and a monophthong [u:], such as in *soul* [sə͡ul]~[suːl] 'soul (music)' and *coach* [kə͡utʂ]~[kuːtʂ] 'id.'. Notably, all three of these possible diphthongs only occur in words of foreign origin, and their phonological status is uncertain.

Despite the scarceness of diphthongs in modern Standard Swedish, three diphthongs, which are represented through various transcriptions such as /ai au øy/ (Bruce, 2010) and /ei au ey/ (Schulte, 2005) due to uncertainty about their phonetic realization (Schulte, 2005), were present in Old Norse and all Nordic varieties derived from it up until about 1000 years ago (Bruce, 2010). At this point, in the middle of the 10th century, a sound change that would eventually lead to the loss of */ai au øy/ in many East Scandinavian varieties started in Denmark. In this so-called East Nordic monophthongization, */ai/ developed into long /e:/ while */au øy/ merged into long /ø:/ (Schulte, 2005). Another important change in the diphthong inventory of Swedish dialects in particular emerged about 500 years ago in southern Sweden, in which phonologically long vowels were diphthongized. A second, separate diphthongization of long vowels also took place in Southern Swedish much later (Bruce, 2010).

The implications of this are that although diphthongs are non-existent or at least uncommon in the standard language, they do appear in several regional varieties of Swedish, both as direct reflexes of the original diphthongs present in Old Norse as well as the result of the later diphthongization processes. Reflexes of */ai au øy/ with retained diphthongal characteristics are commonly referred to in the literature as primary diphthongs, while those stemming from later diphthongization processes are referred to as secondary diphthongs. In some of the present literature, a distinction might be made between secondary diphthongs and diphthongized long vowels (e.g. Bruce, 2010). However, no such distinction will be made here, where any diphthongs which are not direct reflexes of Old Norse diphthongs will be referred to as secondary diphthongs.

Primary diphthongs are retained in some peripheral Swedish dialects, both in Sweden and in Finland. In Sweden, primary diphthongs occur for example in the majority of Jämtland (Dahl et al., 2010), in Västerbotten (Dahl et al., 2010), in Norrbotten (Dahl et al., 2010; Westerlund, 2010), in parts of northern Norrland (Westerlund, 2010), as well as on Gotland (Elert, 1989). In Finland, primary diphthongs are found in most varieties of Finland Swedish, with the exception of the varieties of Swedish spoken on Åland

(Fin. Ahvenanmaa) and in western Nyland (Fin. Uusimaa) (Holm, 1980) where they have been extensively monophthongized. However, the retention of primary diphthongs in a particular dialect does not necessarily mean that all three Old Norse diphthongs have modern-day diphthongal counterparts. While */ai au øy/ remain distinct in some of these areas, for example as /ɛi au oy/ on Gotland, as /ɛi œu œy/ in Liljendal in eastern Nyland (Kim, 2008), */au/ and */øy/ commonly merge in other areas, such as in Rödåliden in Västerbotten where the reflex of */ai/ is /ei/ and the reflex of both */au/ and /øy/ is /øy/ (Westerlund, 2010). In at least one are, namely Nederkalix-Töre in Norrbotten, */ai/ and */øy/ have been monophthongized leaving the reflex of */au/ as the only extant primary diphthong (Holm, 1980).

While secondary diphthongs occur in some of the areas mentioned above, such as in Norrbotten (Westerlund, 2010), northern Norrland (Bruce, 2010), and on Gotland (Elert, 1989; Bruce, 2010) in Sweden, they can also be found in many regions where the primary diphthongs have been monophthongized, such as in Mälardalen in central Sweden, in most parts southern Sweden, and on Gotland (Elert, 1989; Bruce, 2010). In Finland, secondary diphthongs occur for example in parts of Österbotten and Egentliga Finland (Fin. Varsinais-Suomi) (Leinonen, 2007). The most common source of these secondary diphthongs, as has previously been mentioned, is the diphthongization of long vowels. This type of secondary diphthongs occurs extensively in southern Sweden, reaching north up until an imaginary line between the northern border of Halland and Oskarshamn on the east coast (Elert, 1989). While the general characteristics of these secondary diphthongs are similar in all southern Swedish varieties, the first vowel targets as well as the extent of the formant movement of these diphthongs vary (Elert, 1989). Furthermore, while all 9 long vowels are diphthongized in western and southern Skåne, the diphthongization only affects the 4 high vowels /iː yː ʉː uː/ in the Göinge and Kristianstad area and contrastingly only the 6 non-high vowels /eː øː oː ɛː aː/ in Karlskrona in eastern Blekinge (Bruce, 2010). Like in the southern Swedish varieties, all long vowels in the Mälardalen area can be diphthongized, with the non-high vowels becoming centering diphthongs (Elert, 1989; Bruce, 2010) and the high vowels becoming consonantal closing diphthongs, e.g. [ij yj ʉβ uβ] (Bruce, 2010). On Gotland, long /iː yː aː/ are not diphthongized, while /eː ʉː øː uː/ are realized as closing diphthongs and /ɛː oː/ are realized as opening diphthongs, and in northern Norrland the diphthongization commonly affects /eː øː/ and sometimes also /oː ɛː/, which are then realized as closing diphthongs (Bruce, 2010).

2.2.2 Diphthongs in Estonian Swedish

While no previous acoustic studies of Estonian Swedish diphthongs exist, both Danell (1905-34) and Lagman (1979) both feature impressionistic analyses of the diphthongs in the Nuckö dialect. Additionally, Lagman (1979) provides a more general overview of diphthongs in Estonian Swedish as a whole. While there are several rising diphthongs in Estonian Swedish (Lagman, 1979; Danell, 1905-34), they are of little interest here and will not be discussed extensively.

Like in many other peripheral varieties of Swedish, a characteristic feature of all Estonian Swedish dialects is the retention of primary diphthongs (Lagman, 1979; Danell, 1905-34). While the merger of Old Norse */øy au/ into a single diphthong is common in some other Swedish varieties, as described in Section 2.2.1, reducing the total number of primary diphthongs from 3 to 2, Estonian Swedish differs in that it instead features a merger of */ai øy/ into /ai/ in all varieties (Danell, 1905-34), /ai au/ as the two remaining diphthongs. It is possible that the merger of the rounded front diphthong */øy/ with the unrounded front diphthong */ai/ correlates with the unrounding of the rounded front vowels /y ø/ to /i e/ in many of the Estonian Swedish dialects (Lagman, 1979). Due to the amount of variation in the two primary diphthongs across the Estonian Swedish dialects, they will here be described phonemically as /ai au/, although it should be noted that the choice of symbols is rather arbitrary. Additonally, the phonetic transcriptions given below are based on orthographic transcriptions in Lagman (1979) and Danell (1905-34), which means that they at best are rather general approximations.

The two extant primary diphthongs /ai au/ display a considerable amount of variation within Estonian Swedish. For /ai/ Lagman (1979) gives varying realizations such as [a͡i] in Nuckö and Ormsö, [ɛ͡i] in Rågöarna, Nargö, Runö, and Svenskby, and [e͡i] in Vippal and Korkis. These realizations suggest that dialectal variation in /ai/ can be found mainly in the height of the first vowel target. Even more cross-dialectal variation seems to be present in /au/, affecting both the height, backness, and rounding of the first vowel target as well as the backness of the second vowel target. For this diphthong Lagman (1979) reports the realization [a͡u] in Nuckö with the local variants [a͡ʉ] and [ɛ͡ʉ], and Danell (1905-34) further notes that the variation [a͡u] occurs in Österby and Birkas (Est. Pürksi), while [ɛ͡ʉ] is prevalent in Enby (Est. Einbi) (Lagman, 1979). In the other dialects, /au/ occurs as [ɛ͡ʉ] with the local variant [øʉ] in Ormsö, as [ø͡ʉ] in Vippal and Korkis, and as [o͡u] on Runö (Lagman, 1979). Interestingly, judging from forms such as [ɛ͡ig] 'eye' for which Nuckö has /au/ 'eye' in the current data, it seems that /ai/ and /au/ are further merged into a single primary diphthong in Nargö.

In addition to the retained primary diphthongs, several Estonian Swedish dialects have evolved secondary diphthongs from long vowels, as in many other dialects of Swedish. For example, long /oː/ has been diphthongized into a falling diphthong [o͡u] in Nargö and occasionally also in the mainland parts of the Nuckö-speaking area, on Odensholm (Est. Osmussaare), and in Rickul (Est. Riguldi) (Danell, 1905-34), while it conversely occurs as a rising diphthong [u͡a] in Runö and occasionally as [u͡o] in Vippal (Lagman, 1979). A diphthong [ɛ͡i] appears sporadically as a variant of long /ɛ:/ in mainland Nuckö and on Odensholm as well as in Rickul, while also ocurring as an allophone of long /e:/ in mainland Nuckö and Rickul. Furthermore, Danell (1905-34) reports that the diphthong /ɛi/ occurs in some loan words such as /kɛisar/ 'emperor' in Nuckö. However, long vowels are not the only source of secondary diphthongs in the Estonian Swedish dialects. In Nuckö and Rickul, short [ɛ̆͡i] occurs as an allophone of short [ɛ] before velars (Lagman, 1979), and the disappearance of hiatus in the past participle in words such as /tai/ 'taken' and /foi/ 'gotten' in some southern parts of Nuckö, which correspond to /tae/ and /foe/ in other parts of the dialect area, has led to a merger of the two vowels into a single diphthong (Lagman, 1979). A diphthong [a͡ʉ] also occurs in some environments in Vippal-Korkis where the other dialects have /o/ (Lagman, 1979), and is possibly distinct from /au/ which is realized as [ø͡ʉ].

Two secondary diphthongs that are of particular interest in the current study are /oi ui/ which are described to occur in the Nuckö and Ormsö dialects (Lagman, 1979; Danell, 1905-34). Judging from the present data they can also be found in Rickul and Runö, while there are no indications of such diphthongs in Nargö or Rågöarna. These two diphthongs /oi ui/ share a common origin as a result of a sound change involving /v/ plus a following vowel after a syllable-initial /k/ or /h/ (Lagman, 1979). Examples include /huit/ 'white' in Nuckö and Ormsö which corresponds to /kvi:t/ in Dagö, Vippal-Korkis, and Rågöarna, as well as /koin/~/kuin/ 'mill' in Nuckö and Ormsö which corresponds to /kvɛ:rn/ in Vippal-Korkis (Lagman, 1979). While Danell (1905-34) argues that /ui/ in the Nuckö dialect has developed either from earlier /Cwi:/ or a lengthened /wɛ/ preceding the consonant clusters /rn/ and /ld/, the exact details of the evolution of these diphthongs remains unclear. Furthermore, while there are indications both in the literature and in the present data that suggest that /oi/ and /ui/ are in free variation, as is seen in the word /koin/~/kuin/ 'mill' in Nuckö and Ormsö, it is contradicted by the existence of minimal pairs such as /huit/ 'white' and /hoit/ 'wheat', also in the Nuckö (Danell, 1905-34) and Ormsö dialects (Lagman, 1979). However, Danell (1905-34) notes that /oi/ is uncommon in the Nuckö dialect.

A notable characteristic of /ai au oi ui/ which has been described in the literature for the Nuckö dialect (Lagman, 1979; Danell, 1905-34), is that they have short (Lagman, 1979) or reduced (Danell, 1905-34) allophones in some contexts. Lagman (1979) describes the short allophones of /ai oi ui/ as diphthongal [ɛ̆͡i ŏ͡i ŭ͡i], with short /ai/ varying between the diphthong [ɛ̆͡i] and a monophthongal realization [ɛ~e]. Unlike the other diphthongs, short /au/ is exclusively realized as a monophthong [ʉ~ø] (Lagman, 1979). Danell (1905-34) only describes diphthongal short allophones of /ai ui/, which again are [ɛ̆͡i ŭ͡i] respectively; he further notes that the reduced diphthong [ɛ̆͡i] does not occur in the Ormsö dialect, which might either indicate that it has another quality, perhaps [ă͡i], or that it is always realized as a monophthong. It should be noted that the position of the breve in the transcription used here is arbitrarily placed on the first part of the diphthong, but applies to the diphthong as a whole. In their orthographic transcription, Danell (1905-34) and Lagman (1979) make use of an underdot on the second element instead, e.g. <oị> for [ŏ͡i].

The short allophones of /ai au oi ui/ are conditioned by several morphological and phonetic environments. The long diphthongs are commonly reduced in monosyllabic words when they occur as the first part of a compound, e.g. [ho͡it] 'wheat' versus [hŏ͡itsme:ɽ] 'wheat flour' and [ba͡in] 'leg' versus [bɛ̆͡invɛrk] 'leg pain'. This process is not limited to diphthongs, but also affects monophthongs, as is exemplified by [re:] 'red' and [rɛa͡uat] 'red-eyed', and occurs in all Estonian Swedish dialects except for Runö and Nargö (Lagman, 1979). The long diphthongs are also reduced in adjectives taking the neuter singular suffix –t, for example [sa͡in] 'late (predicative feminine singular)' vs [sɛ̆͡int] 'late (predicative neuter singular)' (Danell, 1905-34) and [hu͡it] 'white (predicative feminine singular)' vs [hŭ͡it:] 'white (predicative neuter singular)' (Lagman, 1979), as well as in the past and supine forms of verbs in the 3rd conjugation which take the suffix –t, such as [raike] 'to smoke' vs [rɛ̆͡ift] 'smoked' and [ra͡is] 'to clear' vs [rɛ̆͡ist] 'cleared' (Danell, 1905-34). Reduced diphthongs also turn up in some derived words, such as [gɽɛ̆͡imskur]~[gɽɛmskur] 'forgetful' from [gɽa͡im] 'to forget', and [gɛ̆͡itig:] 'wasp' from [ga͡it] 'goat', and furthermore appear in the passive form of verbs taking the –st suffix as well as in some other environments (Danell, 1905-34), although examples are few in the literature.

2.3 Methods in acoustic diphthong analysis

2.3.1 Analysis based on internal diphthong structure

Based on the notion in earlier literature that the prototypical monophthong is characterized by a vowel target that is acoustically manifested as a relatively stable steady state of the formants in the central portion of the vowel (e.g. Lehiste & Peterson, 1961; Lindblom, 1963), diphthongs have in earlier acoustic studies such as Lehiste and Peterson (1961) and Gay (1968) as well as some more recent studies (e.g. Man, 2007; Roengpitya, 2002; Yu et al., 2004; Hu, 2003; Maxwell & Fletcher, 2010) been investigated and described with reference to an internal structure consisting of clearly identifiable elements that is different to the corresponding structure in monophthongs. The maximal diphthong structure that is present in most of these studies revolves around two vowel targets that each appear as a steady state with little formant movement, which are connected by a transition element during which the formants change from the frequency values of the one target to the frequency values of the other target. Due to variations in terminology and definitions between different authors, these three elements will here be referred to as the first (vocalic) element, the transition, and the second (vocalic) element.

One of the benefits of having this underlying structure is that it allows for the measurement not only of the total diphthong duration, but also the duration of the two vocalic elements and the transition (e.g. Man, 2007; Roengpitya, 2002; Hu, 2003; Tasko & Greilick, 2010; Lehiste & Peterson, 1961; Gay, 1968). These diphthong elements also serve as anchors for spectral measurements, which have commonly been made at the middle point of these elements (e.g. Gay, 1968; Man; 2007; Lehiste & Peterson, 1961; Yu et al., 2004), at the steadiest state in each vocalic element (e.g. Maxwell & Fletcher, 2010), or at the start and end of the transition element (e.g. Tasko & Greilick, 2010). In general, the goal is to place the measurement points where the formants feature the smallest amount of change while at the same time avoiding coarticulatory effects from the surrounding consonants.

A significant drawback of this methodology however is that not all actual diphthongs necessarily fit into this three-element model commonly employed in earlier research. For example, the diphthongs [e͡ɪ] and [o͡ʊ] in American English have been found to feature only a single steady state (Lehiste & Peterson, 1961) or alternatively no real steady states at all (Gay, 1968), while [ɑ͡ɪ ɑ͡ʊ ɔ͡ɪ] display the more canonical three-element structure (Lehiste & Peterson, 1961; Gay, 1968). This reflects one of the central issues in the literature on diphthongs, namely the question whether a diphthong is a single vowel whose quality changes over time or if it is a sequence of two vowels with different qualities that are realized within one syllable (Miret, 1998). Furthermore, it is also possible to question how objective the process of determining the boundaries between the diphthong elements is, although some of the more recent studies have used methods such as defining the beginning and end of the transition as the points where the formant change is less than 20Hz over a time window of 20 ms (Tasko & Greilick, 2010), where the formant change is the lowest (Maxwell & Fletcher, 2010), defining the boundary between the two vocalic elements as the point where the change in formant values reaches its maximum (Asu et al., 2012), or using more intricate definitions such as the zero crossing of the 1st and 2nd derivatives of the F_1 and F_2 (Roengpitya, 2007), rather than relying on visual and auditory judgment alone.

2.3.2 Analysis based on equidistant measurement points

A trend in some recent studies has been to move away from the idea that diphthongs need to be fundamentally different from monophthongs, especially in the light of findings that indicate that monophthongs feature formant change even in the absence of surrounding consonants, and that this spectral variation can be important for the characterization of the vowel (Nearey & Assmann, 1986) as well as being utilized differently in different dialects (Fox & Jacewicz, 2009; Williams & Escudero, 2014). Rather than basing acoustic analysis on smaller elements within both monophthongs and diphthongs, these studies have developed methods that are equally applicable to all vowels. A distinguishing characteristic of these studies is that the formant frequencies are sampled at a number of equidistant time points over the course of the vowel, discarding the need to rely on a predetermined diphthong and monophthong structure. The number of sampling points used range from three points at 25%, 50%, and 75% of the total course of the vowel (Roengpitya, 2002), five time points at 20%, 35%, 50%, 65% and 80% of the total course of the vowel (Mayr & Davies, 2011; Fox & Jacewicz, 2009), nine time points at 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of the total course of the vowel (Asu et al., 2012), or even 30 time points over the central 60% of the vowel (Williams & Escudero, 2014). Besides providing greater objectivity, this method also provides the option of excluding predetermined parts of the vowel that could otherwise contain coarticulatory effects of neighboring consonants, usually the first and last 20% (Asu et al., 2012; Fox & Jacewicz, 2009; Mayr & Davies, 2011; Williams & Escudero, 2014), or 25% (Roengpitya, 2002).

Equidistant measurement points can provide similar measures to those used in earlier studies, such as formant measurements made at the beginning and end of the vowel as well as durational measurements of the total vowel duration as well as duration measurements of parts of the vowel, the latter achieve for example by using the time point with the maximal formant change as the boundary between the first and the second elements (e.g. Asu et al., 2012). However, one of the greatest benefits of this method is the opportunity it provides to employ acoustic measures that are calculated from the raw spectral and temporal data to characterize the dynamic properties of a vowel. Such measures are for example trajectory length and spectral rate of change, which were introduced in Fox and Jacewicz (2009) and later featured in studies such as Mayr and Davies (2011) and Asu et al. (2012).

In Fox and Jacewicz (2009) and Mayr and Davies (2011), trajectory length (TL) is defined as “a measure of formant movement which tracks […] formant frequency change over the course of [a] vowel’s duration” (Fox & Jacewicz, 2009, p. 2606), and represents the total amount of formant change from the first to the last measurement point. It is calculated by first calculating the Euclidean distance between each successive measurement point using the following formula, which gives the trajectory length for each separate vowel section (VSL), i.e. the part of a vowel between two measurement points:

$$VSL_n = \sqrt{(F1_n - F1_{n+1})^2 + (F2_n - F2_{n+1})^2}$$

The total trajectory length of a vowel is then calculated as the sum of all four vowel sections:

$$TL = \sum_{n=1}^{4} VSL_n$$

This measure has been shown to more accurately represent the magnitude of the formant moved compared to vector length, which only represents the Euclidean distance between the first and the last measurement points in the F_1-F_2 vowel space, and thus does not account for changes in the formant movement during the course of the vowel (Fox & Jacewicz, 2009). Notably, trajectory length has been found to vary as a function of dialect (Fox & Jacewicz, 2009; Mayr & Davies, 2011).

Spectral rate of change comes in two varieties in Fox and Jacewicz (2009) and Mayr and Davies (2011). A general spectral rate of change (TL_{roc}) reflects the average speed of the formant change over the central 60% of the vowel, calculated as:

$$TL\_roc = \frac{TL}{0.60 \times v\_dur}$$

The other variety of spectral rate of change is the vowel section spectral rate of change (VSL_{roc}) which represents the average speed of the formant change between each measurement point:

$$VSL\_roc_n = \frac{VSL_n}{0.15 \times v\_dur}$$

Like for TL, both TL_{roc} and VSL_{roc} have been shown to vary as a function of dialect (Fox & Jacewicz, 2009; Mayr & Davies, 2011).

2.3.3 Formant extraction

Regardless of what sampling method is employed, data in the form of formant frequency values must be extracted from the chosen measurement points in order to perform analysis on spectral measures. While this can be done manually, the process of extracting formant data can quickly become impractical and time-consuming as a result of a larger material or when spectral data is sampled from many time points over the course of the vowel, such as when using the equidistant formant measurement method described in Section 2.3.2. It thus becomes evident that an automatic formant frequency measurement method is to be preferred over a manual one in order to speed up the process and reduce the risk of human error. However, as Kabir et al. (2010) put it, “[d]espite numerous attempts to build accurate and reliable automatic formant extractors, there are still no tools available that can automatically extract the “true” formants from the speech in [a] very large corpora” (p. 341).

When working with the software Praat (Boersma & Weenink, 2016), which has been created specifically for phonetic analysis, automatic formant extraction is usually carried out using the Burg algorithm (Anderson, 1978). Using the Burg algorithm in Praat allows for the manipulation of several parameters in order for the analysis to as accurately as possible be able to measure the formants of a given vowel. It is suggested that the frequency of the maximum formant, i.e. the ceiling below which the analysis will look for formants, is set according to whether the speaker is a man, woman, or a child (Boersma & Weenink, 2016). This is done to correct for physiological differences in the vocal tracts of men, women, and children (Escudero et al., 2009), since the Burg algorithm implements an LPC analysis which is based on an approximation of the vocal tract (Bradbury, 2000). The recommended gender-specific values for the formant ceiling in Praat’s formant extraction using the Burg algorithm is 5000 Hz for male speakers and 5500 Hz for female speakers (Boersma & Weenink, 2016). However, when Escudero et al. (2009) investigated these recommended formant ceilings based on F_1 and F_2 values from the central 60% of 5600 vowel tokens in Brazilian and European Portuguese, they found that using fixed gender-specific formant ceilings lead to incorrectly analyzed formant values for some vowels. Some specific problems encountered were the incorrect identification of F_2 as F_1 in back vowels or in the case of /i/ as the second tracheal resonance (Stevens, 1998).

The solution to this problem in Escudero et al. (2009) was to use an optimal formant ceiling procedure where the maximum formant ceiling is increased in small frequency steps between a predetermined low and high frequency value. For each such step, the formants of every individual vowel token are measured. The optimal formant ceiling is then defined as the formant ceiling which results in the lowest variation in the F_1-F_2 pairs for each vowel type, calculated as the variance of the logarithmic F_1 value for that particular vowel type plus the variance of the corresponding logarithmic F_2 value. In Escudero et al. (2009), the maximum formant ceiling was increased in 10 Hz increments between 4500 and 6500 Hz for women and 4000 and 6000 Hz for men, which yielded a total of 201 ceilings for each individual vowel token. The results of this procedure yielded optimal formant ceilings that ranged between 4300 and 5900 Hz for male speakers and 5000 and 6000 Hz for females depending on the vowel category, and by using this method almost all outliers that were present in the original measurements were corrected (Escudero et al., 2009).

This optimal formant ceiling procedure, with variations in the formant ceiling range and size of the formant ceiling increase in each step, has since been used in several other acoustic studies, such as Clason (2009), van Leussen el al. (2011), and Yoon et al. (2015). Notably, Clason (2009) uses this method not only on a single measurement point or measurement area in monophthongs but also on diphthongs using 3 equidistant time points, with a separate optimal ceiling for each time point. While Clason (2009) reports that the procedure still fails to accurately measure the formants in some vowels in some speakers, it features an attractive improvement over relying on the standard suggested values for the Burg algorithm in Praat while still being a relatively simple procedure.

3 Method

3.1 Aims and research questions

As stated in Chapter 1, the aim of present study is to describe the characteristics of the diphthongs /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects in terms of acoustic differences in each diphthong combination, as well as exploring any potential cross-dialectal variation in each of the four diphthongs. This will carried out as an investigation of how several acoustic measures chosen on the basis of recent acoustic studies on diphthongs vary as functions of the factors diphthong category and dialect. The specific research questions that will be targeted in the present study are:

1. How do the diphthongs /ai au oi ui/ differ from each other in terms of their acoustic characteristics in each of the three dialects Nuckö, Ormsö, and Rickul?

2. How does the realization of each of the diphthongs /ai au oi ui/ differ in terms of their acoustic characteristics across dialects?

3.2 Acoustic measures and hypotheses

In order to quantify the acoustic characteristics of the diphthongs, their duration as well as their F_1 and F_2 frequency values at 9 time points over the course of the diphthongs will be measured. The F_3 frequency value is not included in the analysis since the present study does not specifically aim to investigate differences between rounded and unrounded front vowel targets. A realization such as [ø], which has been described for the first target of /au/ in the Ormsö dialect (Lagman, 1979), should still be distinct from a corresponding unrounded vowel by having a lower F_2 value (Lindblad, 2011). Additionally, the exclusion of F_3 enables the results of the current study to be straightforwardly related to the results in earlier studies such as Fox and Jacewicz (2009) and Mayr and Davies (2011), on which a large part of the current methodology has been based. Based on the raw F_1, F_2, and duration data, the analysis will target eight acoustic measures, which can be categorized into three categories: spectral measures, temporal measures, and spectro-temporal measures.

The spectral measures are the initial F_1 (InF_1) and F_2 (InF_2) and the final F_1 ($FinF_1$) and F_2 ($FinF_2$), which correspond to the frequency values of these formants at the 20% and 80% measurement points respectively, and are assumed to roughly represent the characteristics of the diphthongs' first and second vowel targets respectively. While these measures have been analyzed together in the form of F_1-F_2 plots in many earlier studies, they are here investigated separately in the statistical analysis for greater detail and accuracy. With regards to earlier descriptions of Estonian Swedish diphthongs by Lagman (1979) and Danell (1905-34), it is expected to find a certain degree of cross-dialectal variation in the spectral measures of at least /au/, particularly in InF_1, InF_2, and $FinF_2$, reflecting the variation in height, frontness/backness, and rounding of the first vocalic element of this diphthong as well as the varying frontness/backness of the second target, as suggested by the description of various dialectal allophones as [a͡u]~[a͡ʉ]~[ɛ͡ʉ]~[ø͡ʉ]. While Lagman (1979) describes /ai/ as [a͡i] in both Nuckö and Ormsö, it is expected that if dialectal variation does exist in the realization of this diphthong, it will show up in InF_1 and to a lesser degree InF_2, judging from the description that /ai/ occurs varyingly as [a͡i]~[ɛ͡i]~[e͡i] in other Estonian Swedish dialects. Within each dialect, it is expected that /ai/, /oi/, and /ui/ are differentiated by a gradually decreasing InF_1 and InF_2, reflecting the location of the first vowel targets in the vowel space. It is also expected that /ai oi ui/ will have a higher $FinF_2$ than /au/, reflecting their front vs non-front targets, while it is not expected that /ai oi ui/ will differ from each other in terms of $FinF_2$ based on earlier descriptions. Furthermore, it is not expected that $FinF_1$ will vary as a function of either dialect or diphthong type.

The only temporal measure that will be included in this study is the total diphthong duration. This measure has been shown to differ between dialects in English (Fox & Jacewicz, 2009) but not in Welsh (Mayr & Davies, 2011), and while it is not clear from the current knowledge of Estonian Swedish diphthongs how it would vary as a function of either diphthong category or dialect, it is included in the present study to assess this particular question.

Finally, the spectro-temporal measures, which are included on the basis of Fox and Jacewicz (2009), consist of trajectory length (TL), average spectral rate of change (TL_{roc}), and vowel section rate of change (VSL_{roc}), which as described in Section 2.4.2 have previously all been shown to vary as a function of dialect. It is expected that TL could potentially vary both between the different diphthongs due to their differently positioned targets in the vowel space, and also between dialects for at least /au/ and perhaps also /ai/, reflecting cross-dialectal differences in the location of their vowel targets. In the case of TL_{roc}, Gay (1968) suggests that such a measure might be a more stable indicator of differences between diphthong categories such as /aɪ/ and /ɔɪ/ in American English than the F_1 and F_2 frequencies of their targets. However, while Fox and Jacewicz (2009) found that TL_{roc} was not affected by changes in duration that stemmed from varying degrees of emphasis, which further suggests that TL_{roc} might be more independent of changes in duration that TL, they also found that another context that lead to a higher duration, namely the presence of a following voiced plosive, also lead to an increase in TL_{roc}. Thus, while it is expected that TL_{roc} will vary as a function of primarily diphthong category, the results might shed some further light on how it is affected by changes in duration. As for VSL_{roc}, based on the results in Fox and Jacewicz (2009) as well as Mayr and Davies (2011), it is predicted that measure will vary as a function of both dialect and diphthong. In particular, it is expected to reflect the internal structure of the diphthongs, such as the temporal extent of any steady states linked to the vowel targets as well as the extent of the transition.

3.3 Data collection, material, and informants

The data on Estonian Swedish diphthongs that is used in the current study comes from a collection of recordings made specifically for phonetic analysis as a part of the project *Estlandssvenskans språkstruktur* (ESST). These recordings have been collected during several field trips to Stockholm in 2009, 2012, and 2015, as well as to Pürksi (Sv. Birkas) and Einbi (Sv. Enby) in Estonia in 2012. A majority of the Stockholm recordings were made at the Estonian Swedish community *Svenska Odlingens Vänner* (SOV), and the recordings were collected by Susanne Schötz and Eva-Liina Asu, as well as the author of this study during the most recent field trip, using either a Sony portable DAT recorder TCD-D8 or a Roland R-09HR WAVE/MP3 recorder, both with Sony tiepin type condenser microphones ECMT-T140.

The recordings consist of both elicited and read word lists as well as spontaneous dialogues. The word lists aim to elicit certain speech sounds which the phonetic research in the ESST project focused on at the time, such as vowels, liquids, word accents, or sibilants, while the spontaneous dialogues contain conversations between two speakers about a particular topic, such as the informants' childhood in Estonia. Only during the last field trip to Stockholm in 2015 was a word list used that had a dedicated section for eliciting Estonian Swedish diphthongs.

In the current study, all ESST recordings containing diphthongs, both word lists and spontaneous dialogues, were used as a source of diphthong data in order to create as extensive of a dataset as possible. The data extraction process which is described in Section 3.4.1 and 3.4.2 yielded a dataset containing a total of 1072 diphthongs from 26 speakers of 6 different Estonian Swedish dialects, and the distribution of speakers within each dialect in this dataset is shown in Table 1. All speakers were between 64-91 years old at the time of the recordings, with a mean age of 82. It should be noted that some speakers were recorded more than once, so that they appear several times in the current data with different ages. Furthermore, speakers additionally varied in terms of which Estonian Swedish village they originated, which languages they spoke, the amount of time they had spent in Sweden, and their educational background. These factors were not taken into account in the present study.

Table 1. Distribution of speakers per dialect and gender in the initial dataset serving as a basis for the formant normalization procedure. A hyphen (-) indicates a lack of speakers for that particular dialect and gender combination.

	Number of speakers		
Dialect	**Male**	**Female**	**Total**
Nargö	-	1	1
Nuckö	3	2	5
Ormsö	3	7	10
Rickul	2	3	5
Runö	1	2	3
Rågö	2	-	2
Total	**11**	**15**	**26**

3.4 Data preparation

3.4.1 Diphthong segmentation and extraction

All diphthongs present in the recordings made during the field trips described in Section 3.2 were initially assessed according to certain criteria. The first criterion was that each diphthong needed to belong to one of the target diphthong categories /ai au oi ui/. Since variation in the quality of /ai/ and /au/ was expected, these categories were defined as any diphthong starting with a low-to-mid and central-to-front quality and ending in a high front quality and any diphthong ending in a high rounded quality respectively, following Lagman (1979). The second criterion was that any diphthong occurring in interjections, such as /nai/ 'no' and /oi/ 'oh', which were most commonly found in the spontaneous dialogues, were to be excluded. While only a few diphthong tokens appeared in interjections, their exclusion was based on the idea that they could potentially unstable in duration, particularly in comparison with the elicited words from the word lists. The third criterion was that the diphthongs needed to be free of external noise, such as informants speaking at the same or clothes rubbing against the microphone.

The diphthongs that fulfilled the four criteria were manually segmented in Praat (Boersma & Weenink, 2016). The onset of the diphthongs was defined as the appearance of clear formants in the spectrogram and the onset boundary was placed at the zero-crossing of the first positive peak in the waveform belonging to the diphthong. The offset boundary was placed at the zero-crossing of the first positive peak in the waveform belonging to a following voiced segment, at the onset of the occlusion of a following voiceless stop, or at the onset of frication noise of a following fricative. In the case of utterance-final diphthongs, the offset boundary was placed where a significant drop in amplitude of the speech signal could be identified in the waveform, which was usually accompanied by the disappearance of clearly visible formants in the spectrogram. In the few cases where a diphthong would immediately be followed by another vowel, such as in the word /daiur/ 'dies', the end of the diphthong was defined as the point where F_1 and F_2 started to move away from the target position associated with the second vowel target, based on inspection of the spectrogram. An example of the segmentation of the diphthong /ai/ is shown in Figure 2.

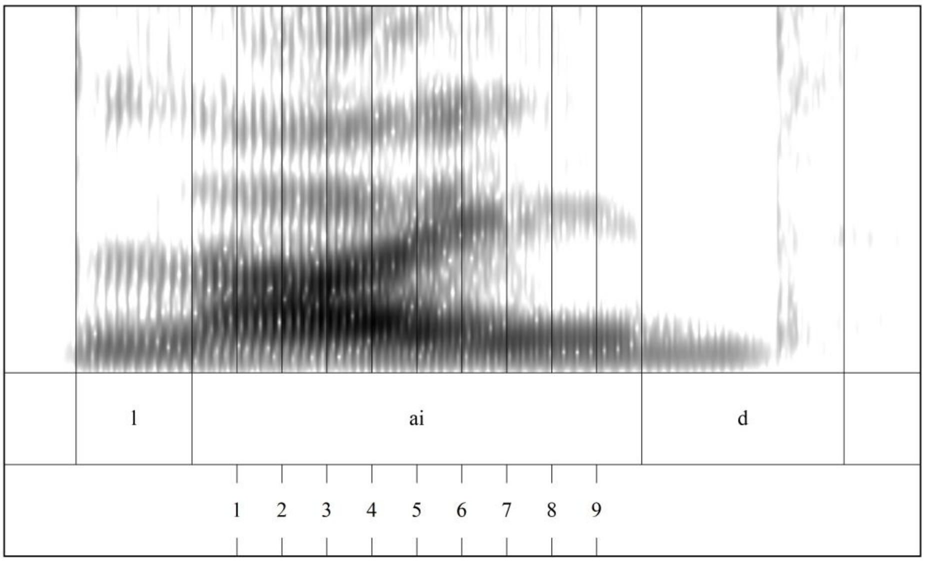

Figure 2. Spectrogram of the word /laid/ '(has) lead' showing the segmentation of /ai/ and the nine equidistant measurement points.

After the segmentation process, all segmented diphthongs in each recording were given a label referring to the diphthong category as well as a number referring to their position relative to the other diphthongs of the same kind in a given recording. These diphthongs were then extracted from their original recordings using a Praat script and saved as individual WAV and TextGrid files. In the extraction process, all diphthong files were given a unique ID based on the name of the recording from which they were extracted together with their original label. This made it possible to easily locate a particular diphthong token in the original recordings if needed, and the names of the recordings additionally contained information about speaker, dialect, and gender.

Following extraction, each diphthong was individually assessed in its original recording and additionally categorized for the factors focus, stress, r-context, and length. Focus was defined as the presence or absence of a clear prominence in the F0 contour, in reference to the Swedish focus accent (Riad, 2014). Stress was defined as the whether the diphthong occurred in a primary stressed, secondary stressed, or unstressed syllable, based on auditory judgement. R-context was defined as the presence or absence of a following /r/ after the target vowel, which could potentially trigger allophonic realizations of some diphthongs (Lagman, 1979). Length was defined as whether a given diphthong occurred in an environment in which diphthong shortening occurs, based on the description of diphthongs and monophthongs in the Nuckö dialect by Lagman (1979) and Danell (1905-34), as presented in Section 2.2.2. In some difficult cases, such as distinguishing /huit/ 'white (predicative feminine singular)" and /huit:/ 'white (predicative neuter singular)', the length of the diphthong could sometimes be judged from the gender of the noun it referred to if such a noun was overt in the recording. In cases where the length of a diphthong could not be reliably determined, it was marked as unknown.

As a final step of the diphthong extraction process, a Praat script was run that added 50 ms of silence at the beginning and end of each sound file, since the Burg algorithm was not able to measure the formants of the diphthongs close to the edges of the sound files due to their short nature.

3.4.2 Formant extraction

In order to capture the dynamic properties of the Estonian Swedish diphthongs, formant measurements were carried out at 9 equidistant sampling points at 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90% of the total diphthong duration, as in Asu et al. (2012), although the analysis would only be based spectral information from measurement point 2 to 8, corresponding to the central 60% of the vowel analyzed in previous studies employing spectro-temporal measures (e.g. Fox & Jacewicz, 2009; Mayr & Davies, 2011). Originally, formant extraction was carried out on each time point in all diphthongs by the means of a Praat script using the Burg algorithm with the suggested maximum formant setting of 5000 Hz for the males and 5500 Hz for the females (Boersma & Weenink, 2016). However, as in the study by Escudero et al. (2009), this resulted in many incorrectly measured formant frequencies. An alternative approach was tested, where the maximum formant value was set manually for each speaker, but the results were again unsatisfactory. Notably, the ideal settings for the earlier time points in a diphthong such as /ai/ were suboptimal for the later time points in the same diphthong and vice versa, due to the large variation in vowel quality within the diphthong itself. In general, the analysis would erroneously identify the second tracheal resonance (Stevens, 1998) as F_2 during the second target of for example /ai/ when the maximum formant frequency was set too low, and when the maximum formant frequency was set too high F_1 and F_2 during the first target of /oi/ and /ui/ were merged into a single formant due to their proximity to each other. A further attempt was made to set individual values for each diphthong and speaker manually, but it was finally decided that an automatic method of optimal formant frequency selection, as suggested by Escudero et al. (2009), would be hugely beneficial. At this point, in order for the automatic optimal ceiling procedure to work, diphthong categories that were only represented by a single token in a given speaker were excluded for that speaker, since the procedure requires at least two tokens per speaker of a given diphthong category in order to calculate the variance of the formants. Additionally, all diphthong files were also manually checked for an unusually high F0 that could potentially disturb the formant analysis, and such files were excluded.

The optimal formant ceiling procedure was initially carried out on each time points in each diphthong category in each speaker using a Praat script with the settings employed in Escudero et al. (2009), namely a formant ceiling increased in steps of 10 Hz between 4000 and 6000 Hz for male speakers and between 4500 and 6500 Hz for female speakers. However, upon inspection of the extracted F_1 and F_2 frequency values for each of the four diphthong categories it was concluded that the procedure once again yielded unsatisfactory results, especially for some of the female speakers. In order to improve the accuracy of the procedure, several other frequency ceiling ranges for the female speakers were tested, among them 5500-8000 Hz and 5000-7000 Hz. The 5000-7000 Hz range was finally chosen as the one that resulted in the least formant measurement errors in the female speakers. The formant extraction process ultimately provided a dataset consisting of F_1 and F_2 measurements at every time point in 1072 diphthongs shown in Table 2, as well as the total duration of these diphthongs.

Table 2. Distribution of diphthong tokens per dialect and diphthong category in the initial dataset serving as a basis for the formant normalization procedure.

	Number of diphthong tokens				
Dialect	**/ai/**	**/au/**	**/oi/**	**/ui/**	**Total**
Nargö	25	-	-	-	25
Nuckö	87	38	13	36	174
Ormsö	213	53	129	70	465
Rickul	162	68	9	21	260
Runö	20	-	12	17	49
Rågö	66	33	-	-	99
Total	**573**	**192**	**163**	**144**	**1072**

3.4.3 Formant normalization

Due to differences in the size and location of the vowel space in male and female speakers stemming from physiological differences in the vocal tracts of the two genders (Flynn et al., 2011; Adank et al., 2004; Thomas & Kendall, 2007), the diphthong data from each gender was originally intended to be analyzed separately. However, due to the limited amount of data for some of the diphthong and dialect combinations it was deemed more advantageous to attempt to reduce gender differences in the formant data, providing more tokens in each level of the two primary factors of interest.

Since the Lobanov method (Lobanov, 1971) has been proven to be one of the best formant normalization methods for removing physiological differences between speakers while retaining sociolinguistic information (Adank et al., 2004), it was chosen to provide normalized formant values that would be comparable across the two genders. The standard Lobanov method is a z-score transformation where the average formant frequency across all vowel categories for a speaker is subtracted from the corresponding formant frequency in a particular vowel token and then divided by the standard derivation for the average formant frequency across all vowel types for that speaker (Adank et al., 2004; Thomas & Kendall, 2007):

$$F_{si}^{Lobanov} = \frac{F_{si} - \mu_{si}}{\delta_{si}}$$

However, a disadvantage of this normalization method is that it is the most successful when all vowels of a speaker's vowel system are included in the normalization process, and can yield skewed results if speakers with differently shaped vowel spaces are included (Thomas & Kendall, 2007). This was judged to be problematic for the current study in that not all of the four diphthongs /ai au oi ui/ occur in each of the 26 speakers, and that a certain amount of dialectal variation is expected in the diphthong formants, especially in /ai/ and /au/. In an attempt to diminish the effects of these known problems, the normalization formula was modified so that instead of using the mean and standard deviation of a given formant across all vowel tokens in a given speaker, the mean and standard deviation of the formant across all vowel types in all speakers of the same gender were used:

$$F_{si}^{Lobanov} = \frac{F_{si} - \mu_{gi}}{\delta_{gi}}$$

This modified normalization procedure was carried out by means of a Praat script on the original F_1 and F_2 values in Hertz. Based on visual investigation of the resulting normalized formant values for the females and male speakers plotted together in an F_1-F_2 diagram, it was deemed that the normalization had been sufficiently successful as the vowel spaces of the two genders largely overlapped, as can be seen in Figure 3. However, it should be noted that the normalized female vowel space displayed slightly more extreme values in the region of cardinal [o] and [u] than the normalized male vowel space, noticeably affecting the InF_2 of /ui/ and /oi/ and also the InF_1 of /ui/.

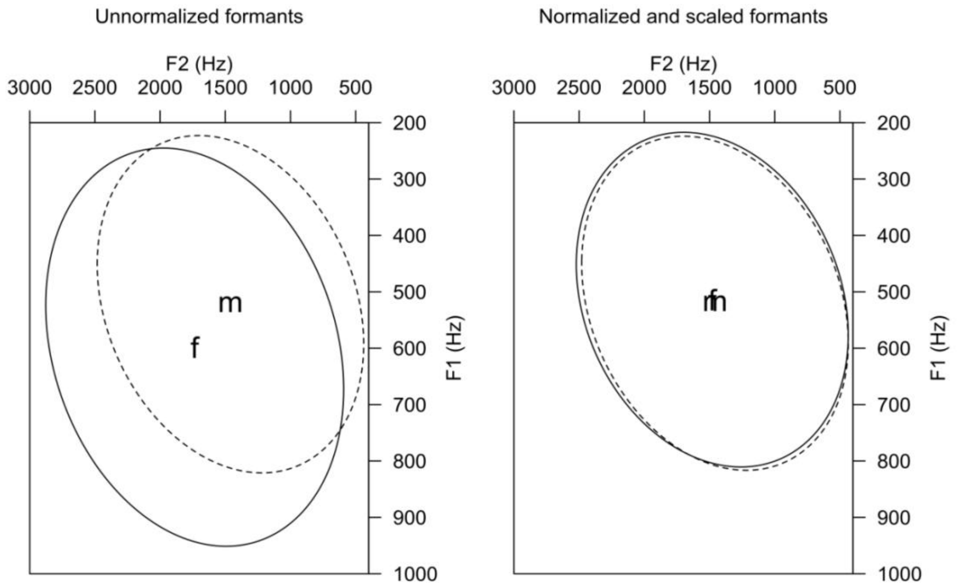

Figure 3. Mean F_1 and F_2 values for all diphthong tokens at all time points in the male (m) and female (f) speakers. Unnormalized formant values are shown to the left, and normalized and scaled formant values are shown to the right. The ellipses correspond to the 95% confidence intervals of the means.

3.4.4 Formant scaling and calculation of spectro-temporal measures

The Lobanov-normalized formant values, which are represented by z-scores placed in the vowel space based on their distance from the mean values of each of the two first formants, do not resemble their unnormalized Hertz counterparts. In order facilitate the analysis, the normalized values were scaled back into Hertz-like values using a Praat script employing an algorithm based on the one suggested by Thomas and Kendall (2007):

$$F_1' = 250 + 500\left(\frac{F_1^N - F_{1MIN}^N}{F_{1MAX}^N - F_{1MIN}^N}\right)$$

$$F_2' = 850 + 1400\left(\frac{F_2^N - F_{2MIN}^N}{F_{2MAX}^N - F_{2MIN}^N}\right)$$

However, by trial and error this original algorithm was modified in such a way that the scaled values for the male speakers more accurately mirror the unnormalized values when plotted in an F_1-F_2 plot:

$$F_1' = 240 + 720\left(\frac{F_1^N - F_{1MIN}^N}{F_{1MAX}^N - F_{1MIN}^N}\right)$$

$$F_2' = 500 + 2300\left(\frac{F_2^N - F_{2MIN}^N}{F_{2MAX}^N - F_{2MIN}^N}\right)$$

The result of this male-based formant scaling is shown in Figure 3. An alternative scaling algorithm was developed based on the formant values for the female speakers, but since there was no need for two alternative scaling algorithms in the actual analysis it was ultimately not used:

$$F_1' = 270 + 860\left(\frac{F_1^N - F_{1MIN}^N}{F_{1MAX}^N - F_{1MIN}^N}\right)$$

$$F_2' = 670 + 2510\left(\frac{F_2^N - F_{2MIN}^N}{F_{2MAX}^N - F_{2MIN}^N}\right)$$

As a final part of the data preparation process, a Praat script was run that provided total diphthong duration, initial F_1 and F_2 values corresponding to the formant frequency values at time point 2, final F_1 and F_2 values corresponding to the formant frequency values at time point 8, as well as the trajectory length, spectral rate of change, and vowel section rate of change for each individual diphthong. All spectral and spectro-temporal measures were calculated both from the unnormalized Hertz values as well as the scaled and normalized formant values. The formula for calculating the trajectory of each individual vowel section (VSL) was the same the one used in Fox and Jacewicz (2009) and Mayr and Davies (2011) described in Section 2.3.2. However, the formula for calculating the total trajectory length (TL) was changed to account both for the larger amount of vowel sections, as well as the exclusion of the first and last two vowel sections lying outside the central 60% of the vowel:

$$TL = \sum_{n=3}^{8} VSL_n$$

The average spectral rate of change (TL_{roc}) was calculated as the TL divided by 60% of the total diphthong duration, using the same formula as in Fox and Jacewicz (2009) and Mayr and Davies (2011) as described in Section 2.3.2. Again due to the greater amount of time points than in earlier studies, the formula for calculating the vowel section rate of change (VSL_{roc}) used in Fox and Jacewicz (2009) and Mayr and Davies (2011) was modified so that each vowel section corresponds to 10% of the total diphthong duration:

$$VSL\_roc_n = \frac{VSL_n}{0.10 \times v\_dur}$$

3.5 Data analysis

In order to target the research question presented in Section 3.1, the original dataset used for formant normalization was subset into a smaller dataset only containing data from the Nuckö, Ormsö, and Rickul dialect. The exclusion of the Nargö, Runö, and Rågö dialects was motivated by the lack of data on some of the diphthongs /ai au oi ui/ in these dialects, as well as a desire to focus the analysis on a smaller group of closely related dialects. In order to further simplify the analysis and avoid the influence of uncontrolled factors, all diphthongs that had been marked as unfocused and unstressed in the segmentation process were excluded from the data, as they were judged to be a potential source of larger variations in both duration and formant range. Furthermore, the analysis was originally planned to additionally investigate the differences between long and short diphthongs, but due to a lack of data on short diphthongs, difficulties in judging the length of some diphthongs, as well as time constraints, only diphthongs that could accurately be identified as belonging to long category were analyzed, and all other diphthongs were excluded. This provided a dataset of 498 diphthong tokens from 20 speakers, which is displayed in Table 3.

Table 3. Distribution of diphthong tokens by dialect and diphthong category in the final dataset used in the analysis.

	Diphthong category				
Dialect	**/ai/**	**/au/**	**/oi/**	**/ui/**	**Total**
Nuckö	62	29	8	13	112
Ormsö	81	32	62	28	203
Rickul	111	63	6	3	183
Total	254	124	76	44	498

The statistical analysis of the acoustic data was carried out in the software R (R Core Team, 2015). The initial analysis was performed as a two-way ANOVA by fitting a linear mixed-effects model to each of acoustic measures except for VSL_{roc}, with the within-subjects factor diphthong category and the between-subjects factor dialect as well as a random effect for speaker using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2013) packages with an alpha level of $\alpha = 0.05$. This general ANOVA gave insight into whether there were any significant main effects of diphthong category and dialect respectively as well as whether any diphthong category*dialect interactions were present in the different acoustic measures. It should be noted that TL, TL_{roc}, and VSL_{roc} were log transformed in order to correct for skewed distributions, while outliers were not removed from the data due to time constraints.

If a significant interaction was present in the ANOVA for a given measure, that particular ANOVA was followed up by post-hoc pairwise comparisons of least-squares means on diphthong pairs within each dialect as well as on dialect pairs within each diphthong category using the lsmeans package (Lenth, 2016) in order to test for simple effects in relation to the two research questions. However, if no interaction was present, one factor was simply averaged across the levels of the other factor. In the case of post-hoc pairwise comparisons, the *p*-values were adjusted for multiple comparisons using the mvt method included in the lsmeans package (Lenth, 2016). For visual analysis, the estimated mean values as well as standard errors of each measure were plotted in interaction plots.

In the case of VSL_{roc}, the ANOVA was performed with the additional within-subjects factor vowel section, which would further reveal a possible main effect of vowel section, as well as diphthong*vowel section, dialect*vowel section, and dialect*diphthong*vowel section interactions. The post-hoc pairwise tests were here carried out in the same way as for the other acoustic measures, but on each vowel section separately and adjusted for a total of 180 comparisons.

4 Results

4.1 Spectral characteristics

4.1.1 General spectral characteristics

As a general overview of the spectral characteristics of the diphthongs /ai au oi ui/, their average initial and final F_1 and F_2 frequency values corresponding to the approximate locations of their first and second vowel targets are plotted by dialect in Figure 4 together with the individual means for each diphthong category and speaker. As is evident from Figure 4, the first vowel target of /ai/ occupies the central bottommost portion of the vowel space in all three dialects, and although there is some within-dialect variation in its quality and also some variation in height between dialects, it generally appears to correspond to a low vowel [a]. The second target of /ai/ displays quite some variation in height between speakers in all three dialects, and the relative position of this target compared to the first target of /ui/ suggest that it is broadly located in the area of [i]~[ɪ]~[e] in all three dialects. In the case of /oi/, the spectral behavior of its first target suggests that there is a certain degree of inter-dialectal variation in height in the [o]~[ɔ] area, with the first target being higher in Ormsö and lower in Nuckö. As for the second target of /oi/, it seems to be located mainly in the area of [ɪ], although one speaker in Ormsö and one speaker in Nuckö seem to feature a more back realization which can be approximately described as [ɨ]. The first target of /ui/ seems to display little cross-dialectal variation, suggesting [u] as a plausible realization in all three dialects. In the same way, the second target does not vary considerably between dialects, and its location in the vowel space suggests that it is [i]~[ɪ] in all dialects.

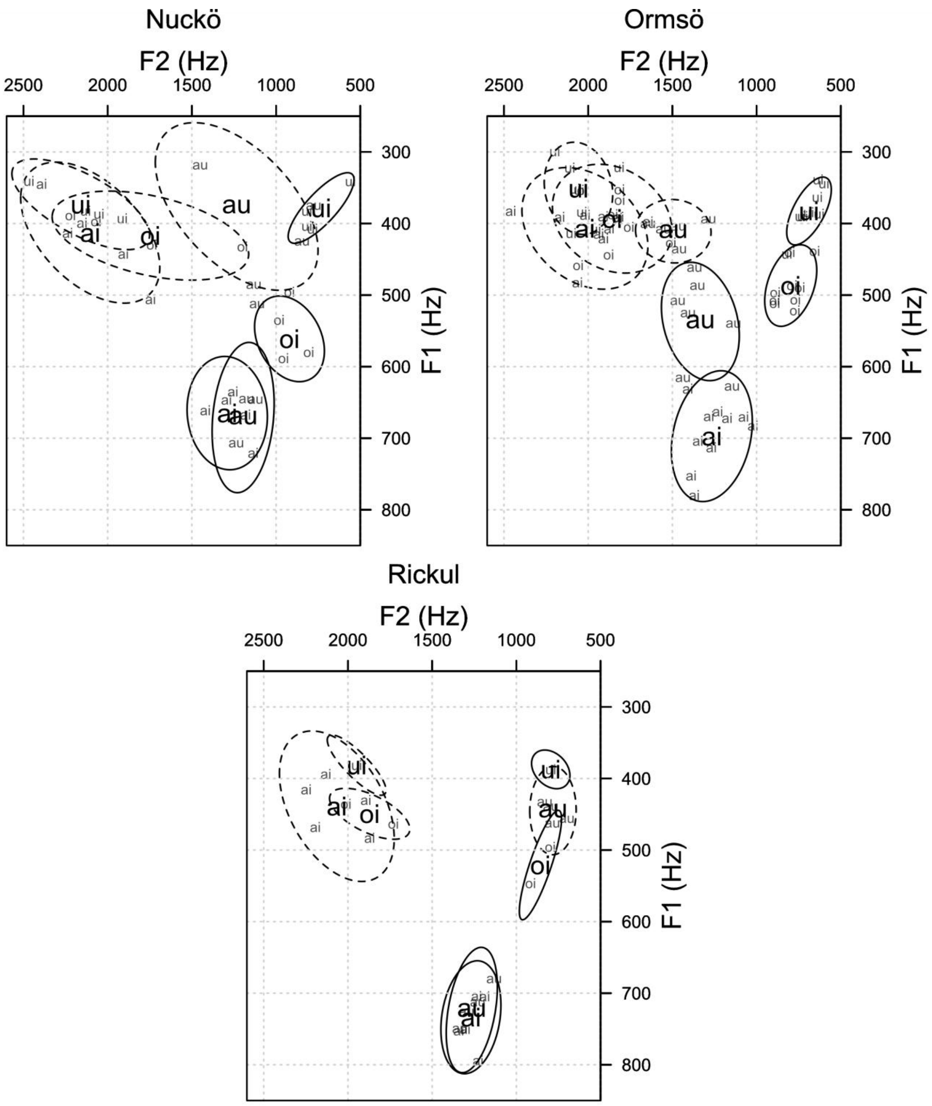

Figure 4. Mean F_1 and F_2 frequency values of the first and second vowel targets of /ai au oi ui/ in each dialect, collected at time point 2 and 8. The ellipses correspond to ±1 standard deviation from the bivariate mean. Each smaller symbol represents the mean for that diphthong in a single speaker.

Unlike in /ai oi ui/, first target of /au/ displays considerable differences between dialects, primarily between Ormsö on the one hand and Nuckö and Rickul on the other. The first target of Rickul /au/ is virtually identical to that of /ai/, suggesting [a] as its only realization. The same mostly holds true for Nuckö, although one speaker features a considerably higher and slightly more back first target of /au/; auditory inspection suggests that its quality is in fact closer to [ɞ]. In the case of Ormsö, the first vowel target of /au/ is considerably higher than that of /ai/ for most speakers, suggesting [œ] as a plausible main realization, although the realization in two of speakers suggests a quality closer to [a] and yet another speaker displays a more back realization, for which [ʌ] is suggested with the aid of auditory analysis. In a similar way, the second target of /au/ displays a considerable amount of variation both between and within dialects. Its location suggests a somewhat lowered but not centralized [u̞] in Rickul, a quality in the area of [ʏ] or [ʉ̞] in Ormsö, and a larger range of realizations [u̞]~[ɵ]~[ʉ] in Nuckö with substantial variation between speakers.

Summarizing these results, the visual analysis of the spectral location of the first and second targets of /ai au oi ui/ indicates possible realizations of /ai/ as [a͡i]~[a͡ɪ]~[a͡e] in all three dialects, of /oi/ as [o͡ɪ]~[o͡ɨ] in Ormsö, [ɔ͡ɪ]~[ɔ͡ɨ] in Nuckö, and [o͡ɪ]~[ɔ͡ɪ] in Rickul, of /ui/ as [u͡ɪ]~[u͡i] in all three dialects, and of /au/ as [a͡u̯] in Rickul, [œ͡ʉ̯]~[œ͡ʏ]~[ʌ͡ʏ]~[a͡ʉ̯] in Ormsö, and [a͡u̯]~[a͡ɵ]~[a͡ʉ]~[ɘ͡u̯] in Nuckö. Figure 5 additionally displays the mean trajectories of the diphthongs in each dialect, and it is evident that that all diphthongs, perhaps with the exception of Rickul /ui/, have mostly unidirectional trajectories which indicate linear formant change from one vowel target to the other without any signs of triphthongization. In the following sections, each acoustic measure will be investigated in further detail, and the general patterns observed here will be discussed again in Chapter 5 in the light of the statistical analysis.

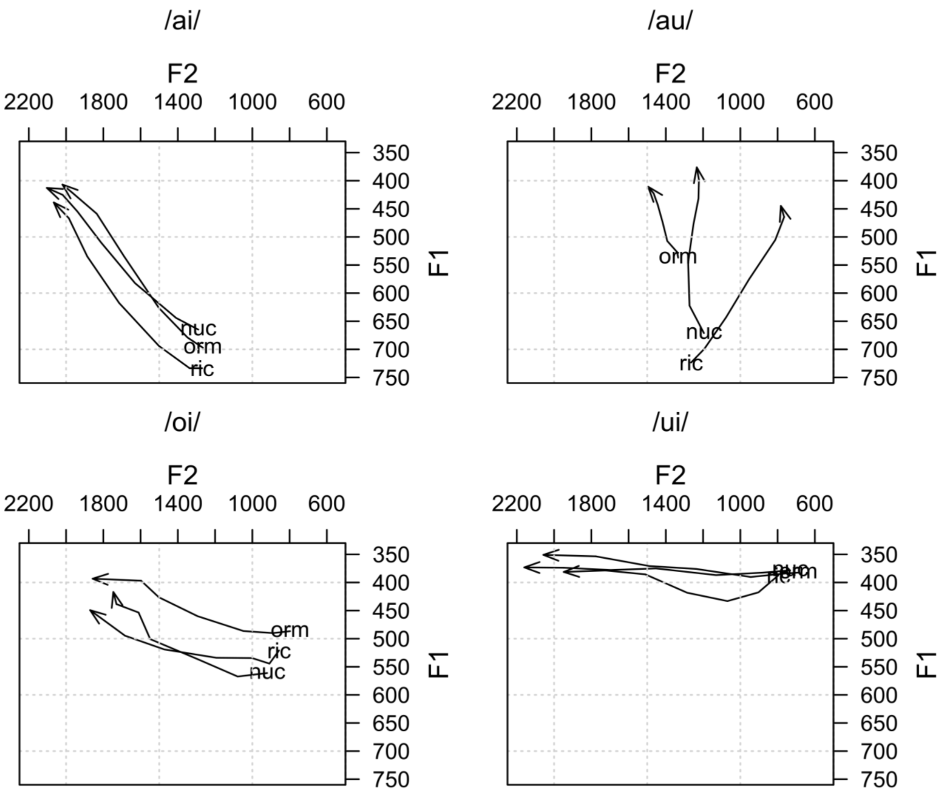

Figure 5. Mean formant trajectories of the diphthongs /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects, refkecting the mean F_1 and F_2 frequency value at each time point from 2 to 8.

4.1.2 Initial F_1

The linear mixed-models ANOVA revealed significant main effects of both diphthong category ($F(3, 475.37) = 359.49$, $p < .001$) and dialect ($F(2, 19.97) = 6.67$, $p = .006$) on the initial F_1 (InF_1), as well as a significant diphthong category*dialect interaction ($F(6, 481.97) = 3.73$, $p = .001$).

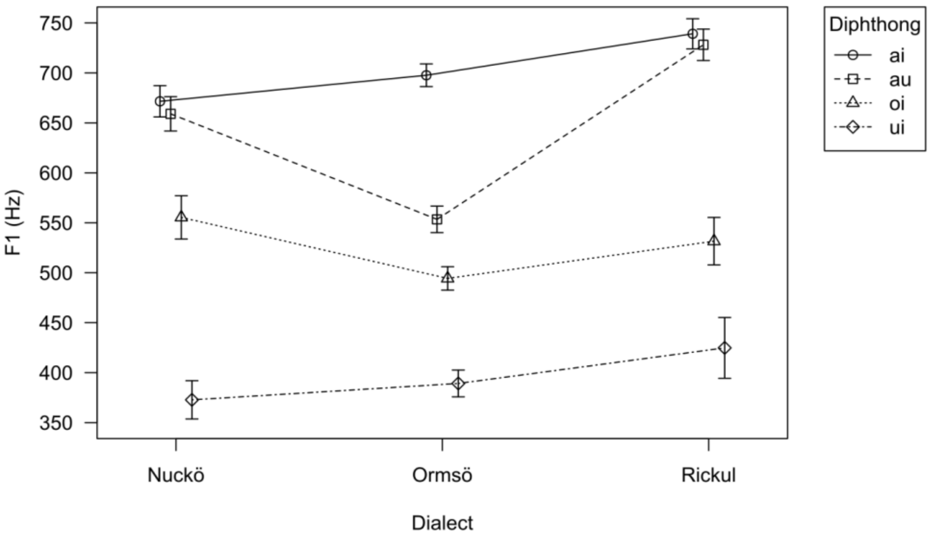

Figure 6. Estimated mean initial F_1 frequency (InF_1) in Hertz (Hz) for /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects. The error bars represent ±1 standard error of the mean, and overlap indicates the presence of a non-significant difference.

Figure 6 suggests clear differences between several of the diphthong categories in each dialect, and investigation of within-dialect differences between diphthong categories through pairwise comparisons of least-squares means, the results of which can be seen in Table 4, showed that /oi ui/ have a lower InF_1 than /ai au/ in all three dialects, and that /ui/ furthermore has a lower InF_1 than /oi/. Contrastingly, the relation of /ai/ and /au/ was found to vary between dialects, where /au/ has a lower InF_1 than /ai/ in Ormsö but not in Nuckö and Rickul. As is evident from Table 5, the cross-dialectal comparison of each diphthong category further revealed that Ormsö /au/ has a lower InF_1 than /au/ in the Nuckö and Rickul dialects. However, although Figure 5 suggests that Nuckö /oi/ has a higher InF_1 than Ormsö /oi/ and that Rickul /ai/ has a higher InF_1 than Nuckö /ai/, no significant differences between dialects were found in the other three diphthongs. The main effect of dialect which was revealed in the ANOVA was not investigated further.

Table 4. Results of the post-hoc pairwise comparisons of the initial F_1 (InF_1) in diphthong pairs within each dialect. Significant results at the multiple comparisons corrected α-level of 0.05 are represented by their *p*-values, while non-significant results are marked with a hyphen (-).

	Dialect		
Diphthong pairs	**Nuckö**	**Ormsö**	**Rickul**
/ai/-/au/	-	$p < .001$	-
/ai/-/oi/	$p < .001$	$p < .001$	$p < .001$
/ai/-/ui/	$p < .001$	$p < .001$	$p < .001$
/au/-/oi/	$p < .001$	$p < .001$	$p < .001$
/au/-/ui/	$p < .001$	$p < .001$	$p < .001$
/oi/-/ui/	$p < .001$	$p < .001$	$p = .023$

Table 5. Results of the post-hoc pairwise comparisons of the initial F_1 (InF_1) in dialect pairs within each diphthong category. Significant results at the multiple comparisons corrected α-level of 0.05 are represented by their *p*-values, while non-significant results are marked with a hyphen (-).

	Diphthong category			
Dialect pairs	**/ai/**	**/au/**	**/oi/**	**/ui/**
Nuckö-Ormsö	-	$p = .001$	-	-
Nuckö-Rickul	-	-	-	-
Ormsö-Rickul	-	$p < .001$	-	-

4.1.3 Initial F_2

The linear mixed-models ANOVA on the initial F_2 (InF_2) measure revealed a significant main effect of diphthong category ($F(3, 475.38) = 242.95$, $p < .001$) as well as a significant diphthong category*dialect interaction ($F(6, 475.58) = 8.3$, $p < .001$), but no significant main effect of dialect ($F(2, 20.79) = 0.03$, p > .05).

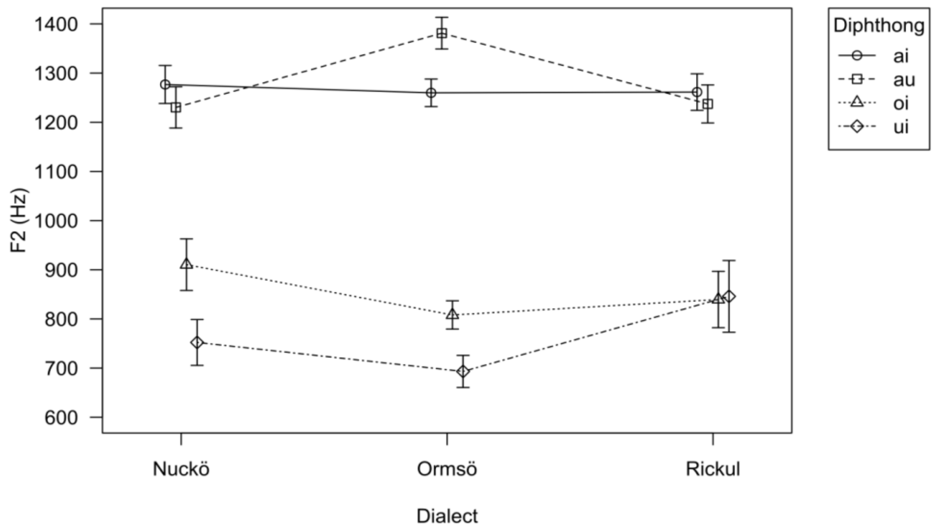

Figure 7. Estimated mean initial $\mathbf{F_2}$ frequency (In$\mathbf{F_2}$) in Hertz (Hz) for /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects. The error bars represent ±1 standard error of the mean, and overlap indicates the presence of a non-significant difference.

Table 6 shows the results of the subsequent post-hoc tests, which indicated that /ai au/ have a consistently higher InF_2 than /oi /ui/ in all dialects, which can be observed in Figure 7 as a clear split between the two groups of diphthongs. Furthermore, smaller differences between the dialects in the within-dialect diphthong systems were revealed, in that /au/ was found to have a higher InF_2 than /ai/ in Ormsö but not in Nuckö or Rickul, and that /ui/ has a lower InF_2 than /oi/ in Nuckö and Ormsö but not in Rickul. Despite these results, the pairwise comparison of each diphthong across the three dialects did not reveal any significant cross-dialectal variation in any of the diphthong categories.

Table 6. Results of the post-hoc pairwise comparisons of the initial F_2 (InF_2) in diphthong pairs within each dialect. Significant results at the multiple comparisons corrected α-level of 0.05 are represented by their *p*-values, while non-significant results are marked with a hyphen (-).

	Dialect		
Diphthong pairs	**Nuckö**	**Ormsö**	**Rickul**
/ai/-/au/	-	$p < .001$	-
/ai/-/oi/	$p < .001$	$p < .001$	$p < .001$
/ai/-/ui/	$p < .001$	$p < .001$	$p < .001$
/au/-/oi/	$p < .001$	$p < .001$	$p < .001$
/au/-/ui/	$p < .001$	$p < .001$	$p < .001$
/oi/-/ui/	$p = .031$	$p < .001$	-

4.1.4 Final F_1

For the final F_1 ($FinF_1$), the linear mixed models-ANOVA revealed a significant main effect of diphthong category ($F(3, 477.24) = 5.11$, $p = .002$), but a non-significant main effect of dialect ($F(2, 21.84) = 2.25$, $p > .05$) and a non-significant diphthong category*dialect interaction ($F(6, 477.43) = .934$, $p > .05$).

Due to the absence of interactions in the ANOVA, post-hoc pairwise comparisons were only carried out on diphthong pairs, with the $FinF_1$ averaged over the levels of the dialect factor. This analysis revealed that across all three dialects /ai au oi/ are not different from each other in terms of $FinF_1$ and form a single grouping with a high $FinF_1$, while /ui/ has a significantly lower $FinF_1$ than /ai/ ($p = .001$), /au/ ($p = .002$), and /oi/ ($p = .004$), a pattern which can also be observed in Figure 8.

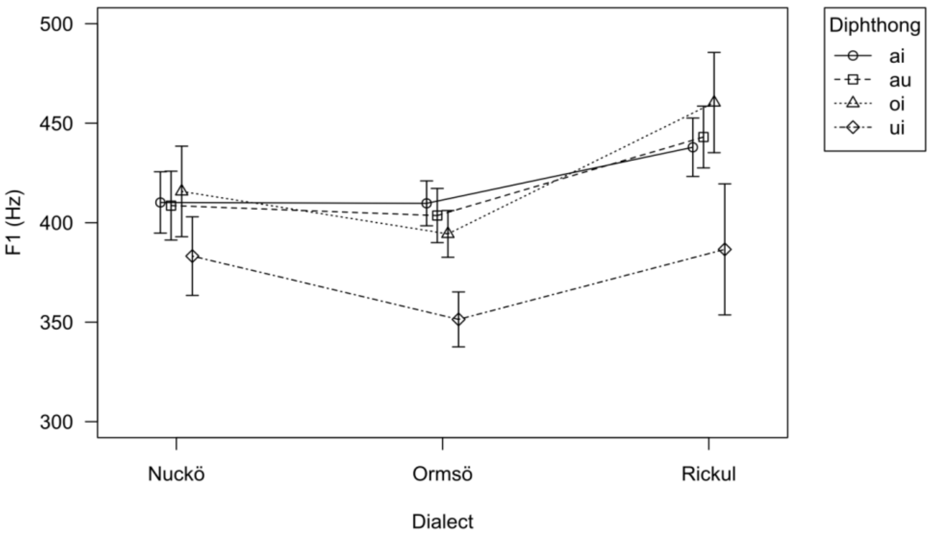

Figure 8. Estimated mean final $\mathbf{F_1}$ frequency (Fin$\mathbf{F_1}$) in Hertz (Hz) for /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects. The error bars represent ±1 standard error of the mean, and overlap indicates the presence of a non-significant difference.

4.1.5 Final F_2

The linear mixed-models ANOVA revealed a significant main effect of diphthong category ($F(3, 473.53) = 698.2$, $p < .001$) as well as a significant diphthong category*dialect interaction ($F(6, 473.68) = 55.04$, $p < .001$) on the final F_2 (FinF_2), but a non-significant main effect of dialect ($F(2, 17.95) = 1.45$, $p = .261$).

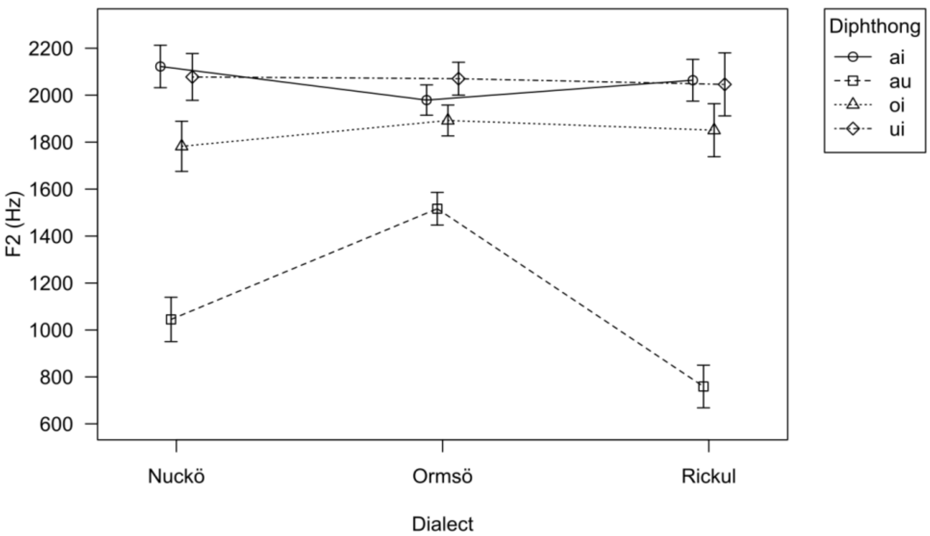

Figure 9. Estimated mean final $\mathbf{F_2}$ frequency (Fin$\mathbf{F_2}$) in Hertz (Hz) for /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects. The error bars represent ±1 standard error of the mean, and overlap indicates the presence of a non-significant difference.

Further analysis through pairwise comparisons, the results of which are shown in Table 7, showed that /ai/ and /ui/ consistently form a single grouping with the highest $FinF_2$ in all dialects. Variation between the dialects in the diphthong systems was found in the relation between /oi/ and /ai ui/; Nuckö /oi/ has a significantly lower $FinF_2$ than both /ai/ and /ui/ while Ormsö /oi/ has a significantly lower $FinF_2$ than /ui/ but not /ai/, and Rickul /oi/ is not significantly different from any of the two diphthongs /ai ui/. As is clearly evident from Figure 9, /au/ has a consistently lower $FinF_2$ than /ai oi ui/ in all three dialects, and as noted in Section 4.1.1 the second element of /au/ displays prevalent dialectal variation. As evidence of this variation, the post-hoc pairwise tests revealed that Ormsö /au/ has a higher $FinF_2$ than /au/ in Nuckö and Rickul, as seen in Table 8. However, although Nuckö /au/ displays a higher $FinF_2$ than Rickul /au/ in Figure 9, the difference was not found to be significant, and neither were any other significant differences found between dialects.

Table 7. Results of the post-hoc pairwise comparisons of the final F_2 ($FinF_2$) in diphthong pairs within each dialect. Significant results at the multiple comparisons corrected α-level of 0.05 are represented by their *p*-values, while non-significant results are marked with a hyphen (-).

	Dialect		
Diphthong pairs	**Nuckö**	**Ormsö**	**Rickul**
/ai/-/au/	$p < .001$	$p < .001$	$p < .001$
/ai/-/oi/	$p < .001$	-	-
/ai/-/ui/	-	-	-
/au/-/oi/	$p < .001$	$p < .001$	$p < .001$
/au/-/ui/	$p < .001$	$p < .001$	$p < .001$
/oi/-/ui/	$p = .005$	$p < .001$	-

Table 8. Results of the post-hoc pairwise comparisons of the final F_2 ($FinF_1$) in dialect pairs within each diphthong category. Significant results at the multiple comparisons corrected α-level of 0.05 are represented by their *p*-values, while non-significant results are marked with a hyphen (-).

	Diphthong category			
Dialect pairs	**/ai/**	**/au/**	**/oi/**	**/ui/**
Nuckö-Ormsö	-	$p = .01$	-	-
Nuckö-Rickul	-	-	-	-
Ormsö-Rickul	-	$p < .001$	-	-

4.2 Temporal characteristics

4.2.1 Duration

While the linear mixed-model ANOVA revealed a significant main effect of diphthong category ($F(3, 482.24) = 3.93$, $p = .009$) as well as a significant diphthong category*dialect interaction ($F(6, 481.97) = 3.73$, $p = .001$) on duration together with a non-significant main effect of dialect ($F(2, 29.21) = 0.47$, $p > .05$), Figure 10 indicates that duration displays a less coherent behavior across dialects than the spectral measures described in Section 4.1.

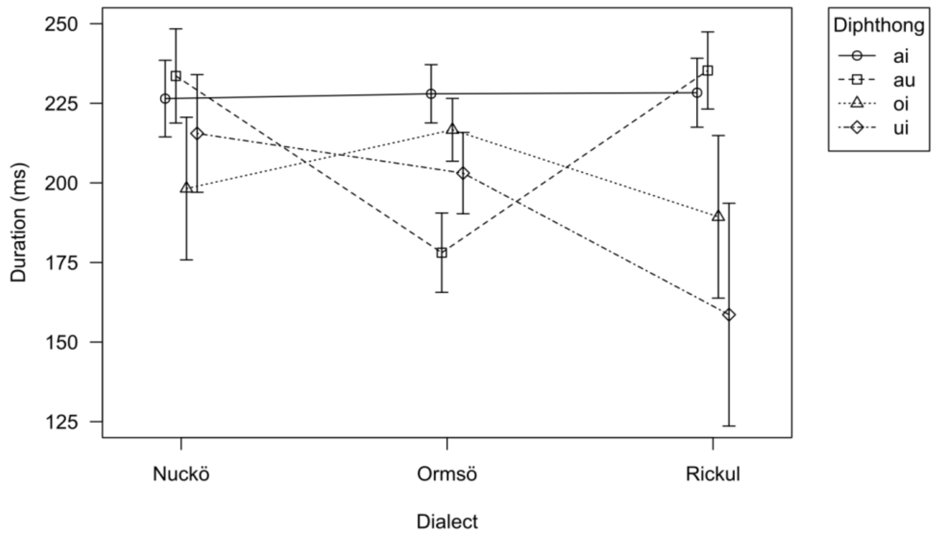

Figure 10. Estimated mean total diphthong duration in milliseconds (ms) for /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects. The error bars represent ±1 standard error of the mean, and overlap indicates the presence of a non-significant difference.

The post-hoc analysis showed that /ai au oi ui/ are not significantly different from each other in terms of duration in Nuckö and Rickul. However, a significant difference was found between /au/ and /ai/ in Ormsö, where /au/ has a lower duration than /ai/ (p = .002), but neither of the two were different from /oi ui/. Furthermore, the post-hoc pairwise tests targeting inter-dialectal differences in the realization of each diphthong additionally showed a marginally significant difference between Ormsö and Rickul /au/ (p = .043), while no significant inter-dialectal differences were found in the other diphthongs. In addition to these results, it is noteworthy that Figure 10 displays a range in duration for some of the diphthong and dialect combinations, such as /oi/ in Nuckö and /oi ui/ in Rickul.

4.3 Spectro-temporal characteristics

4.3.1 Trajectory length

For the spectro-temporal measure trajectory length (TL), the linear mixed-model ANOVA revealed a significant main effect of diphthong category ($F(3, 474.61) = 198.65$, $p < .001$) as well as a significant diphthong category*dialect interaction ($F(6, 474.83) = 11.94$, p < .001), but no significant main effect of dialect ($F(2, 19.12) = 0.34$, $p = .718$).

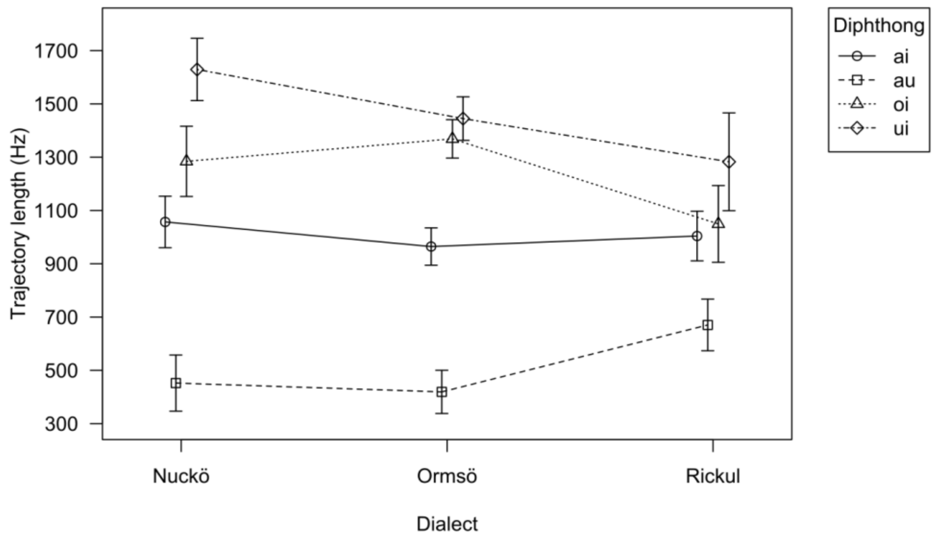

Figure 11. Estimated mean trajectory length (TL) in Hertz (Hz) between time point 2 and 8 for /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects. The error bars represent ±1 standard error of the mean, and overlap indicates the presence of a non-significant difference.

Subsequent analysis of factor levels showed that /au/ has a noticeably shorter TL than /ai oi ui/ in all three dialects, which is evident from Figure 11. The relationship between /oi/ and /ui/ also remains cross-dialectally stable in that there is no significant difference between the two diphthongs in any dialect. However, there is variation across all three dialects in how /ai/ relates to /oi ui/, which is evident from Table 9. In Ormsö, /ai/ has a shorter TL than both /oi/ and /ui/, while it in Nuckö has a shorter TL than /ui/ but not /oi/, and in Rickul there is no significant difference between any of the three diphthongs /ai oi ui/. In terms of cross-dialectal differences in the realization of diphthong categories, Rickul /au/ was found to have a longer TL than /au/ in Ormsö (p = .004), and although Figure 11 shows that the mean TL of Nuckö /au/ is closer to that of /au/ in Ormsö than in Rickul, no significant difference was found between Nuckö and Rickul in this diphthong. The pairwise comparisons revealed no other indications of cross-dialectal variation in TL in the other diphthong categories. However, as an additional general observation, it is notable in Figure 10 that the range of diphthong means for TL is considerably smaller in Rickul than in Nuckö and Ormsö.

Table 9. Results of the post-hoc pairwise comparisons of the trajectory length (TL) in diphthong pairs within each dialect. Significant results at the multiple comparisons corrected α-level of 0.05 are represented by their p-values, while non-significant results are marked with a hyphen (-).

	Dialect		
Diphthong pairs	**Nuckö**	**Ormsö**	**Rickul**
/ai/-/au/	p < .001	p < .001	p < .001
/ai/-/oi/	-	p < .001	-
/ai/-/ui/	p < .001	p < .001	-
/au/-/oi/	p < .001	p < .001	p = .001
/au/-/ui/	p < .001	p < .001	p < .001
/oi/-/ui/	-	-	-

4.3.2 Average spectral rate of change

The results for the average spectral rate of change (TL_{roc}) measure, shown in Figure 12, were found to be similar to those described for TL in Section 4.3.1. The linear mixed-model ANOVA revealed a significant main effect of diphthong category ($F(3, 475.34) = 182.18$, $p < .001$) on TL_{roc} and a significant diphthong category*dialect interaction ($F(6, 475.56) = 3.65$, $p = .001$), but no significant main effect of dialect ($F(2,19.78) = 0.41$, $p > .05$).

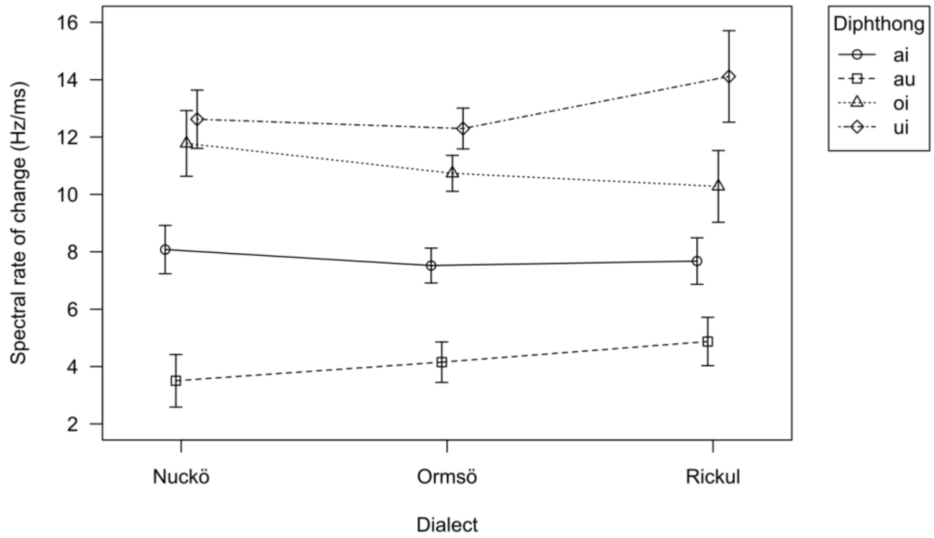

Figure 12. Estimated mean spectral rate of change (**TL_{roc}**) in Hertz per millisecond (Hz/ms) over the central 60% of /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects. The error bars represent ±1 standard error of the mean, and overlap indicates the presence of a non-significant difference.

Like for TL, the post-hoc analysis of differences between diphthong categories, summarized in Table 10, found that the TL_{roc} of /au/ is lower than that of /ai oi ui/ in all three dialects and that /ai/ has a consistently lower TL_{roc} than /ui/, while /oi ui/ display a consistently non-significant difference in TL_{roc} across dialects. Variation between dialects is here found in the relation between /ai/ and /oi/, where /ai/ has a lower TL_{roc} than /oi/ in Nuckö and Ormsö but not in Rickul. However, unlike TL, TL_{roc} was not found to display any significant cross-dialectal differences in the four diphthong categories.

Table 10. Results of the post-hoc pairwise comparisons of the spectral rate of change (TL_{roc}) in diphthong pairs within each dialect. Significant results at the multiple comparisons corrected α-level of 0.05 are represented by their p-values, while non-significant results are marked with a hyphen (-).

	Dialect		
Diphthong pairs	**Nuckö**	**Ormsö**	**Rickul**
/ai/-/au/	$p < .001$	$p < .001$	$p < .001$
/ai/-/oi/	$p = .009$	$p < .001$	-
/ai/-/ui/	$p < .001$	$p < .001$	$p = .001$
/au/-/oi/	$p < .001$	$p < .001$	$p < .001$
/au/-/ui/	$p < .001$	$p < .001$	$p < .001$
/oi/-/ui/	-	-	-

4.3.3 Vowel section rate of change

As is evident from Figure 13, the general shape of the vowel section rate of change (VSL_{roc}) contour of each diphthong category is largely similar across all three dialects. /ai/ stands out compared to /au oi ui/ by having a distinct early peak in the VSL_{roc} contour around vowel section 4 in Nuckö and vowel section 5 in Ormsö and Rickul, before gradually decreasing back to a lower value. This is contrasted by /au/, which overall has a low rate of change in all vowel sections and changes little across the duration of the diphthong; only in Rickul does a somewhat distinct peak emerge around vowel section 6. The VSL_{roc} contour of /ui/ does not show any signs of a peaking behavior, although rather than having a flat contour like /au/ its VSL_{roc} increases until vowel section 5 in all dialects from where it lies relatively stable at a high value until the end of the diphthong, with the initial increase being the steepest in Nuckö and Ormsö but relatively gradual in Rickul. Figure 13 suggests that the most cross-dialectal variation in the shape of the VSL_{roc} contour is found in /oi/, in which the VSL_{roc} increases quickly to peak at a high value at vowel section 6 in Nuckö and Rickul from where it decreases to a lower value, especially in Nuckö. This peaking behavior is not prevalent in Ormsö, where /oi/ instead closely mirrors the VSL_{roc} contour of /ui/, albeit at a lower rate of change in vowel sections 5-7.

The linear mixed-model ANOVA, which was here carried out with the within-subjects factors diphthong and vowel section and the between-subjects factor dialect, revealed significant main effects of diphthong category ($F(3, 2868.34) = 120.73$, $p < .001$) and vowel section ($F(5, 2897.90) = 30.71$, $p = < .001$) on VSL_{roc} as well as significant diphthong category*dialect ($F(6, 2858.71) = 2.536$, $p = 0.019$), diphthong category*vowel section ($F(15, 2897.90) = 4.09$, $p < .001$), dialect*vowel section ($F(10, 2897.90) = 4.47$, $p < .001$), and diphthong category*dialect*vowel section ($F(30, 2897.90) = 2.6$, $p < .001$) interactions. However, the main effect of dialect ($F(2, 21.38) = 1.16$, $p = .332$) on VSL_{roc} was not found to be significant.

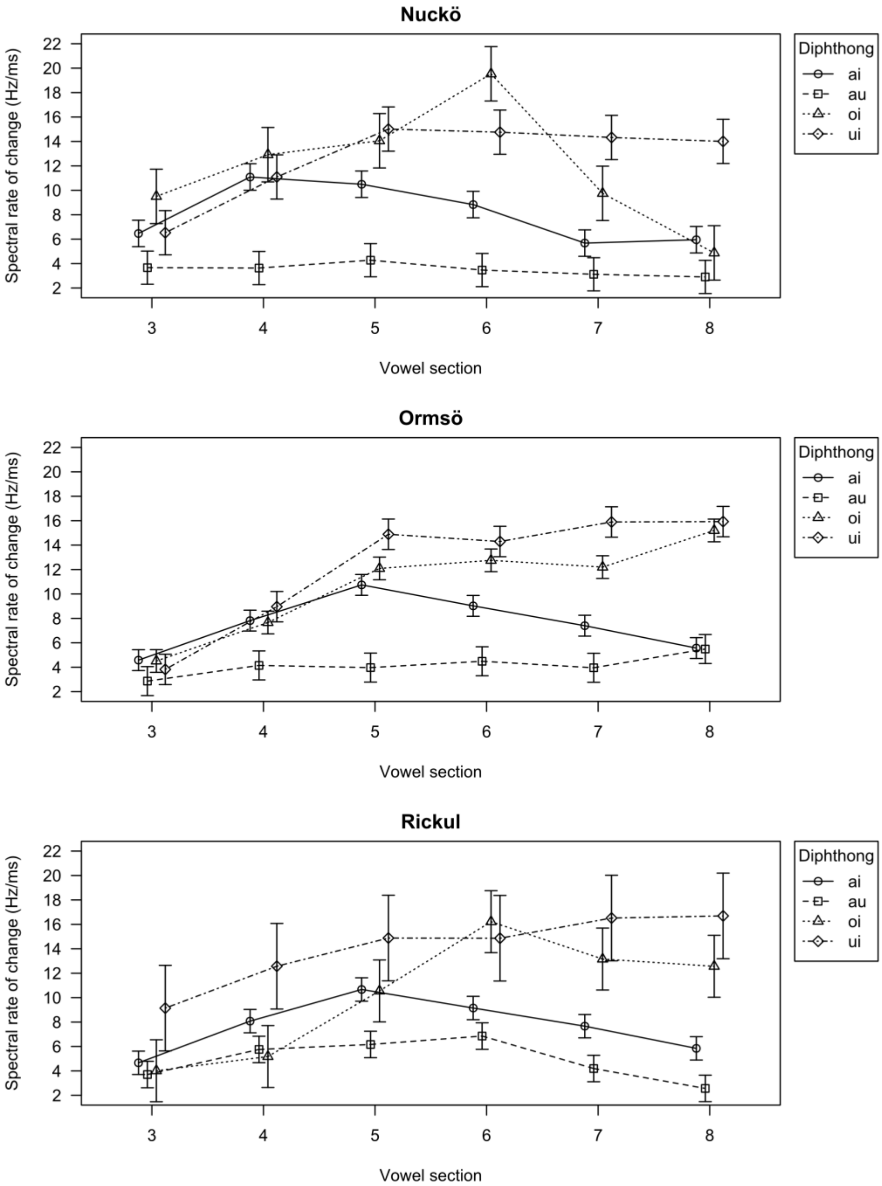

Figure 13. Estimated mean spectral rate of change in each vowel section (VSL_{roc}) in Hertz per millisecond (Hz/ms) over the central 60% of /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects. The error bars represent ±1 standard error of the mean, and overlap indicates the presence of a non-significant difference.

Post-hoc pairwise comparisons of least-squares means of diphthong category within dialect and dialect within diphthong category were made for each vowel section, and the *p*-values were adjusted for all 180 comparisons using the mvt method (Lenth, 2016). These pairwise comparisons, the results of which are shown in Table 11, revealed that a difference between /ai oi ui/ and /au/ was prevalent in all three dialects. In the /ai/-/au/ pair, /ai/ showed a higher VSL_{roc} in vowel section 4-6 in Nuckö, in vowel section 3-7 in Ormsö, and in vowel section 5, 7 and 8 in Rickul. The comparison of the /au/-/oi/ pair showed that /oi/ has a higher VSL_{roc} than /au/ in vowel section 4-6 in Nuckö, in vowel section 3-8 in Ormsö, and in vowel section 7 and 8 in Rickul. Finally, it was revealed that /ui/ has a higher VSL_{roc} than /au/ in vowel section 5-7 in Nuckö, in vowel section 4-8 in Ormsö, and in vowel section 8 in Rickul.

Table 11. Results of the post-hoc pairwise comparisons of the spectral rate of change in vowel section 3-8 (VSL_{roc}) in diphthong pairs within each vowel section and dialect. Significant results at the multiple comparisons corrected α-level of 0.05 are represented by their *p*-values, while non-significant results are marked with a hyphen (-).

		Vowel section					
Dialect	**Diphthong pairs**	**3**	**4**	**5**	**6**	**7**	**8**
Nuckö	/ai/-/au/	-	$p < .001$	$p < .001$	$p = .001$	-	-
	/ai/-/oi/	-	-	-	-	-	-
	/ai/-/ui/	-	-	-	-	$p < .001$	-
	/au/-/oi/	-	$p = .008$	p = .032	$p < .001$	-	-
	/au/-/ui/	-	-	$p < .001$	$p < .001$	$p < .001$	-
	/oi/-/ui/	-	-	-	-	-	-
Ormsö	/ai/-/au/	$p = .009$	$p = .004$	$p < .001$	$p < .001$	$p = .002$	-
	/ai/-/oi/	-	-	-	-	$p = .005$	$p < .001$
	/ai/-/ui/	-	-	-	-	$p = .006$	$p < .001$
	/au/-/oi/	$p = .038$	$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p = .019$
	/au/-/ui/	-	$p = .001$	$p < .001$	$p < .001$	$p < .001$	$p = .007$
	/oi/-/ui/	-	-	-	-	-	-
Rickul	/ai/-/au/	-	-	$p < .001$	-	$p < .001$	$p < .001$
	/ai/-/oi/	-	-	-	-	-	-
	/ai/-/ui/	-	-	-	-	-	-
	/au/-/oi/	-	-	-	-	$p = .027$	$p < .001$
	/au/-/ui/	-	-	-	-	-	$p = .002$
	/oi/-/ui/	-	-	-	-	-	-

While /ai oi ui/ and /au/ were revealed to be significantly different in at least one vowel section in all three dialects, the differentiation of the pairs /ai/-/oi/ and /ai/-/ui/ varies between the dialects. In Rickul, the pairwise comparisons revealed no significant differences between these two pairs in any vowel section. In Nuckö, no significant difference was found in the /ai/-/oi/ pair, but /ui/ was found to have a higher VSL_{roc} than /ai/ in vowel section 7. In line with the results for the pairs /ai/-/au/, /au/-/oi/, and /au/-/ui/, Ormsö also featured the greatest differentiation of the /ai/-/oi/ and /ai/-/au/ pairs out of the three dialects where /ai/ was found to have a lower VSL_{roc} than /oi/ but a higher VSL_{roc} than /ui/ on vowel section 7 and 8. For the last diphthong pair /oi/-/ui/, no significant differences were found in any of the three dialects. The pairwise comparisons of diphthong realizations across dialects further revealed that the only significant cross-dialectal difference in VSL_{roc} occured between Ormsö and Rickul in vowel section 8 in /au/, where Ormsö /au/ has a significantly higher VSL_{roc} than Rickul /au/ (p = 0.35), although the difference is only marginally significant.

5 Discussion

The aims of the current study were to investigate how the diphthong categories /ai au oi ui/ relate to each other within the Nuckö, Ormsö, and Rickul varieties of Estonian Swedish in terms of their acoustic characteristics, as well as to explore potential cross-dialectal variation in these diphthongs by comparing their realizations across the three dialects. The characteristics of the diphthongs were operationalized as eight acoustic measures, reflecting both the static as well as the dynamic aspects of the diphthongs. Generally, in line with previous impressionistic descriptions (Lagman, 1979; Danell, 1905-34), all four diphthongs categories /ai au oi ui/ were found to be acoustically different from each other in all three dialects. Any two diphthongs categories were differentiated from each other by at least two of the acoustic measures, and all eight acoustic measures, although arguably only marginally duration, were found to be used in the characterization of the diphthong categories. Additionally, significant cross-dialectal variation was found in /au/ in four out of the eight acoustic measures.

Rather than solely relying on visual analysis of the approximate locations of the F_1 and F_2 frequency values of the diphthong targets, the current study included the initial F_1 and F_2 (InF_1 and InF_2) and the final F_1 and F_2 ($FinF_1$ and $FinF_2$) frequency values as separate spectral measures in order to statistically test potential differences in the first and second vowel targets of the diphthongs. In the light of the statistical analysis of these measurements, in conjunction with the visual analysis of the position of the diphthong targets in the F_1-F_2 vowel space, it is concluded that the main realizations of /ai/, /oi/, and /ui/ are [a͡ɪ]~[a͡e], [o͡ɪ]~[ɔ͡ɪ], and [u͡i] respectively in all dialects. While some cross-dialectal variation in the distinction of /oi/ from /ai ui/ was found in $FinF_2$ in the statistical analysis, it is argued here that realizations of /ai oi ui/ do not vary considerably between the dialects. The reasoning behind this is that the variation observed between diphthong categories in $FinF_2$ can be argued to stem from the small number of tokens for Nuckö /oi/, and the deviant $FinF_2$ values for two speakers in Nuckö and Ormsö observed in Section 4.1.1 where the second target of /oi/ is closer to [ɨ] than to a front vowel are possibly indicative of a known measurement error where the second tracheal resonance is interpreted as the F_2 (Escudero et al., 2009; Clason, 2009). In the same way, the non-significant difference between Rickul /oi/ and /ui/ in the InF_2 measure, a difference which was found to be significant in the other dialects, can be argued to be caused by the fact that Rickul /ui/ is only represented by 3 tokens from a single male speaker, since the vowel spaces for the male and female speakers do not overlap perfectly in the high back region of the vowel space, as described in Section 3.4.3. It is also possible that such variation between genders causes some of the variation observed in the height of the first target of /oi/, which in the light of the statistical analysis lies more or less in the [o]~[ɔ] region for all three dialects. Furthermore, neither of these differences in the within-dialect diphthong systems were found to give rise to significant-cross dialectal differences in the statistical analysis, which might further suggest that they are relatively minor, and in general supports the interpretation that /ai oi ui/ display similar characteristics across dialects. Thus, as initially hypothesized, it is argued that /ai oi ui/ are differentiated by a gradually rising InF_1 and InF_2 in all three dialects, while the FinF2 measure does not differentiate between /ai oi ui/ but instead serves to distinguish these three front-closing diphthongs from /au/. Additionally, there are no strong signs of potential cross-dialectal variation in /ai/ in any of the measures, while contrary to the initial expectations the $FinF_1$ measure displays unexpected variation between diphthong categories where the second target of /ui/ is located higher in the vowel space than the second target of /ai oi/ across all three dialects. This statistical result governed the transcription of the final target of /ai oi/ primarily as [ɪ] but the final target of /ui/ as [i], which was not immediately apparent from the visual analysis alone. It is speculated that this difference is connected to the fact that the trajectory of /ui/ already has a high quality at the first vowel target, and perhaps also that the formant change in /ui/ mainly occurs along the F_2 axis, thus not limiting $FinF_1$ through target undershoot in the F_1 dimension.

While no significant signs of inter-dialectal variation were found in the realization of /ai oi ui/, the results for InF_1, InF_2, and $FinF_2$ corroborate the hypothesis that /au/ displays considerable dialectal variation in the Nuckö-Ormsö-Rickul dialect area, which is mainly expressed as variation in the height of the first target and the backness of the second target, as initially expected. Although significant cross-dialectal differences in /au/ were not found between Nuckö and Rickul, the general spectral results in Section 4.1.1 suggest that Nuckö /au/ is indeed different from Rickul /au/ in that its second target varies across a large portion of the vowel space from a fully back [u̯] to a central [ʉ], while the second target of Rickul /au/ is a stable [u̯] in all speakers. Thus, it is here suggested the main realizations of /au/ in Nuckö and Ormsö are [a͡u̯]~[a͡ʉ] and [œ͡ʏ] respectively, which generally agrees with the descriptions of these diphthongs found in Danell (1905-34) and Lagman (1979), although the second target of Ormsö /au/ seems to be a more fronted [ʏ] for most speakers rather than [ʉ] as suggested by the orthographic transcription <öu> in Lagman (1979). It should also be noted that while potential variants such as [a͡ɵ]~[ɜ͡u̯] and [ʌ͡ʏ]~[a͡ʉ̯] were identified in some speakers in Nuckö and Ormsö respectively, the realization [ɛ͡ʉ] which Lagman (1979) describes as a possible variant of /au/ in the Nuckö and Ormsö dialects was not found in the current material. Additionally, while the realization of Rickul /au/ is not described in the literature, there is clear evidence that its main realization is [a͡u̟] in all Rickul speakers included in the present study.

The three spectro-temporal measures trajectory length (TL), average spectral rate of change (TL_{roc}), and vowel section rate of change (VSL_{roc}) were included in the present study on the basis of two earlier studies of dialectal differences between monophthongs and diphthongs (Fox & Jacewicz, 2009; Mayr & Davies, 2011) to further characterize differences both between dialects and also potentially between diphthong categories. Like InF_1 and $FinF_2$, the results show that TL reflects cross-dialectal differences in /au/. However, unlike the results for the two spectral measures, the results for TL point at Rickul as the deviant dialect rather than Ormsö, in that Rickul /au/ displays a longer trajectory than Ormsö /au/, for which the difference is significant, and quite possibly also Nuckö /au/. This difference likely stems from the fully back second target in Rickul [a͡u̟], which features movement in both the F_1 and the F_2 dimension, while the fronted second target in Nuckö and Ormsö /au/ leads to formant movement primarily in the F_1 dimension, which has less influence on TL. In addition to cross-dialectal differences, TL also indicates differences in formant movement between diphthong categories, although variation in the relation between /ai/, /oi/ and /ui/ across dialects, possibly due to few tokens in /oi/ and /ui in the Rickul and Nuckö dialects or variation in duration, means that only /au/ on the one hand and /ai oi ui/ on the other hand are consistently differentiated by TL across all dialects in the statistical analysis. In this regard, TL_{roc} appears to be a more successful indicator of diphthong category than TL, differentiating three groups /au/, /ai/, and /oi ui/ in Nuckö and Ormsö, while /ai/ and /oi ui/ still remain indistinct in Rickul. The observation that TL_{roc} might be a more stable indicator of diphthong category than TL potentially agrees with the observation in Gay (1968) that the formant rate of change is more stable across varying degrees of speaking rates than spectral measurements such as vowel target locations. However, contrary to TL, TL_{roc} did not reflect the previously observed inter-dialectal differences in /au/.

In addition to the average rate of formant change represented by the TL_{roc} measure, VSL_{roc} proved to be rather successful in providing detailed information about the location of the greatest formant change in the diphthongs. All four diphthongs proved to be cross-dialectally similar except for a difference between Ormsö and Rickul /au/ in vowel section 8, although the location of this difference suggests that it might very well be due to differences in consonant context, and the results indicate that the diphthongs could potentially be categorized according to the shape of their VSL_{roc} contour: early peak for /ai/, late peak or late plateau for /oi/, late plateau for /ui/, and low level for /au/. This result possibly indicates two distinct steady states in /ai/ where the second one is longer, two distinct steady states in /oi/ where the first one is longer or alternatively only one initial steady state, only one initial steady state in /ui/, and a relative smooth but slow change throughout /au/. Additionally, the difference between the front-closing diphthongs /ai oi ui/ and the back- or central-closing diphthong /au/, which can potentially be manifested in most vowel sections as seen in Ormsö, is the most widespread pattern across the dialects and generally reflects the differentiation between these diphthong categories found in $FinF_2$ and TL. Further differentiation between the fronting diphthongs, found most prevalently in Ormsö but also to a degree in Nuckö, only seems to occur in later vowel sections, and it is hypothesized that the non-significant differences between the fronting diphthongs in Rickul is due to the large variation in /ui/ and /oi/, which can be observed in Figure 13, again possibly stemming from a small amount of tokens for these two diphthong categories in the dialect.

As the sole temporal measure, total diphthong duration was included in the study in order to assess its usefulness in quantifying differences between diphthong categories as well as differences between dialects. However, the difference between Ormsö /au/ and /ai/ is the only significant difference found between diphthong categories, and the difference between Ormsö and Rickul in /au/ is the only significant cross-dialectal difference. While it is possible to argue that the shorter duration of Ormsö /au/, both in relation to Ormsö /ai/ and to Rickul /au/, is correlated with a shorter distance between its vowel targets in this diphthong compared to its counterpart in the other dialects, thus decreasing the amount of time needed to reach its second target, this interpretation is discouraged by the fact that Ormsö /au/ and Nuckö /au/ have similar TLs but highly different durations. Another possible source of the deviant behavior of Ormsö /au/ could be the misclassification of some short instances of Ormsö /au/ as long in the transcription process, but this argument is contradicted by the fact that no short diphthongal realization of /au/ is described in Lagman (1979) and Danell (1905-34), making such misclassifications unlikely for this particular diphthong. Thus, the most likely reason for the durational behavior of Ormsö /au/ is argued to be the effect of some other factor which was not controlled for, perhaps consonant context, and based on the current results it is difficult to assess whether or not true dialectal differences exist in the duration of the Estonian Swedish diphthongs.

While the status of duration as an indicator of differences between levels of the diphthong category and dialect factors remains unclear, it is possible that duration might play a bigger role as an additional influencing factor on the spectral and spectro-temporal measurements. Subtle variation between dialects in the distinction of some of the diphthong categories in the spectral measures has already been discussed as potentially having its source in the small amount of tokens for Rickul and Nuckö /oi/ and for Rickul /ui/, and additionally also in possible measurement errors in the formant extraction process. In a similar way, it is possible to argue that the variation across dialects that can be observed in the relationship between /ai/, /oi/, and /ui/ in TL and TL_{roc} also arises from these sources. However, in the light of the temporal results, an alternative hypothesis can be proposed where the observed cross-dialectal variation in the TL and TL_{roc} of /oi/ and /ui/ is correlated with their durational behavior. The most striking piece of suggestive evidence of such a correlation is the behavior of Rickul /oi/ and /ui/, which display both a lower duration and shorter TL, and for Rickul /ui/ also a higher TL_{roc}, than their Nuckö and Ormsö counterparts. As already discussed, faster speaking rates have been found to reduce the amount of formant movement in American English diphthongs in the form of second target undershoot (Gay, 1968), and it can thus be hypothesized that a shorter duration in the Estonian Swedish diphthongs would either give rise to a higher TL_{roc} in order for the articulators to have time to reach their targets, or a shorter TL if the articulators fail to do so, or possibly a combination of the two. Variation in duration could then potentially explain the subtle observed cross-dialectal variation in measures such as TL, TL_{roc}, and also $FinF_1$. Furthermore, the large variation in the duration of Rickul /oi/ and /ui/ could also serve as an alternate explanation as to why the Rickul diphthongs are not as well distinguished in terms of VSL_{roc} compared to the other dialects; a large variation in duration leads to a high variation in other acoustic measures. However, although this explanation is compelling in that it serves to explain at least a part of the small inter-dialectal differences in the way the within-dialect diphthong systems are organized, care should be taken when speculating about the relationship between duration, TL, and TL_{roc}, which has previously been found to be complex and involve many different factors (Fox & Jacewicz, 2009).

As it stands, it is evident that the current study faces certain limitations. A small sample size with few tokens for some diphthong and dialect combinations limits not only the power of the statistical analysis, but also the accuracy of the formant extraction process leading to possible measurement errors, and ultimately the certainty of the results. However, the incorporation of such a larger dataset was not achievable at the time of this study due to the limited amount of recordings available, particularly taking into consideration the small size of the extant Estonian Swedish community. It should also be noted that while the current study set out to explore variation in the Nuckö-Ormsö-Rickul dialect area by treating the three dialects as homogeneous in the statistical analysis due to limitations of data availability and time, it is evident from previous literature (Lagman, 1967; Danell, 1905-34) as well as from the results presented in Section 4.1.1 that a considerable within-dialect variation in the diphthongs exists, particularly in the case of /au/. Thus, while the current study proved successful in identifying some of the dialectal variation present in the Nuckö-Ormsö-Rickul dialect area, it presents a somewhat generalized picture of this variation. Furthermore, it can also be argued that the current study is limited by certain methodological choices; it is unknown to what extent the modified Lobanov procedure can be seen as a valid alternative to the original normalization formula (Lobanov, 1971; Thomas & Kendall, 2007), and although the normalization procedure was judged to be largely successful, it possibly introduced a degree of variation in the first and second targets of /ui/ and /oi/ as well as in the second target of /ai oi ui/

since the high front and high back areas of the male and female vowel spaces did not match up perfectly. However, again due to the limitations of a small sample size, it was in this particular study judged to be more beneficial than harmful to employ this normalization procedure in order to eliminate the gender factor rather than analyzing each gender separately. It has also become evident, as has already been discussed, that a control for additional factors such as duration and also consonant context would have been beneficial in the interpretation of the spectral and particularly the spectro-temporal measures, although the inclusion of such additional factors were beyond the scope of this study. Despite these limitations, the present study features a great improvement over previous impressionistic descriptions by providing a first acoustic analysis of the diphthongs /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects. Through the use of testable acoustic measures, notably including spectral measures reflecting the location of the vowel targets of the diphthongs as well as spectro-temporal measures reflecting the diphthongs' dynamic characteristics, it has allowed for a more robust picture of the relations in the within-dialects diphthong systems as well as the variation between these dialects. The current study has thus laid a foundation for future research on Estonian Swedish diphthongs, and its methodology can be applied to diphthongs and monophthongs alike, in any language.

6 Conclusion

The present study constitutes the first acoustic study on Estonian Swedish diphthongs, and aimed at establishing the distinguishing characteristics between the four diphthongs categories /ai au oi ui/ in the Nuckö, Ormsö, and Rickul dialects along with investigating potential cross-dialectal variation in each of these diphthong categories. Based on data sampled across the central 60% portion of 498 diphthongs from 20 speakers, the analysis was carried out on eight spectral, temporal, and spectro-temporal acoustic measures, which together quantified both the static and the properties of the diphthongs.

In line with earlier impressionistic descriptions (Lagman, 1979; Danell, 1905-34), the findings indicate that /ai oi ui/ are similarly realized as [a͡ɪ]~[a͡e], [o͡ɪ]~[ɔ͡ɪ], and [u͡i] in all three dialects, and their characteristics are primarily reflected as differences in the F_1 frequency of the first target (InF_1), the F_2 frequency of the first target (InF_2), and the F_1 frequency of the second target ($FinF_1$), while the F_2 frequency of the second target ($FinF_2$) primarily differentiates between the front-closing /ai oi ui/ and the back- or central-closing /au/. In contrast to /ai oi ui/, though agreeing previous descriptions such as Lagman (1979), /au/ displays considerable cross-dialectal variation with the main realizations [a͡ʊ̣] in Rickul, [a͡ʊ̣]~[a͡ʉ] in Nuckö, and [œ͡ʏ] in Ormsö, where the influence of the dialect factor is particularly prevalent in the height of the first target and the backness of the second target, which is reflected as cross-dialectal differences in the InF_1 and $FinF_2$ measures. Additionally, diphthong identity, particularly the distinction between the front-closing diphthongs /ai oi ui/ and the back- or central-closing diphthong /au/ but also to a degree the distinction between /ai/ and /oi ui/, was reflected in the spectro-temporal measures trajectory length (TL), average spectral rate of change (TL_{roc}), and the spectral rate of change of each vowel section (VSL_{roc}). In the case of TL, cross-dialectal differences in /au/ were also apparent, primarily between the Rickul and Ormsö dialects.

Despite limitations such as a small sample size, a previously untested normalization method, and a lack of control for factors such as variation in duration and consonant context and the disregard for intra-dialectal variation in the statistical analysis, the current study contributes considerably to our understanding of the Estonian Swedish diphthongs /ai au oi ui/ in that it for the first time has given an overview of the acoustic characteristics of these diphthongs and their cross-dialectal variation, lending support to many of the observations in earlier impressionistic descriptions (Lagman, 1979; Danell, 1905-34). Furthermore, the methodology and results presented here, along with the limitations of the study, set the stage for future investigations of Estonian Swedish vowels. Future studies should seek to investigate the effect of duration on the acoustic measurements employed here, particularly in the light of the contrast between long and short diphthongs in Estonian Swedish, which would undoubtedly lead to interesting results and increase our understanding of how these measures can serve to characterize vowels in general. The present study has also made it apparent that there is much more to learn about the dialectal variation in the Estonian Swedish diphthongs, and suggests that such variation is likely to be found on a sub-dialectal level, particularly in /ai/ and /au/.

References

Adank, P., Smits, R. & van Hout, R. (2004). A Comparison of Vowel Normalisation Procedures for Language Variation Research. *Journal of the Acoustical Society of America, 116*(5). 3099-3107.

Anderson, N. (1978). On the calculation of filter coefficients for maximum entropy spectral analysis. In Childers: *Modern Spectrum Analysis* (pp. 165-172). IEEE Press.

Árnason, K. (2011). *The Phonology of Icelandic and Faroese*. Oxford University Press.

Asu, E. L., Ewald, O. & Schötz, S. (2015b). The realisation of sj- and tj-sounds in Estonian Swedish : some preliminary results. In Svensson Lundmark, M., Ambrazaitis, G. & Weijer, J. v. d. (Eds.) *Working Papers 55* (pp. 23-26). Lund University.

Asu, E. L., Lippus, P., Niit, E. & Türk, H. (2012). The acoustic characteristics of monophthongs and diphthongs in the Kihnu variety of Estonian. In *Linguistica Uralica, XLVIII*(3), pp. 161–170.

Asu, E. L., Nolan, F. & Schötz, S. (2015a). Comparative Study of Estonian Swedish Voiceless Laterals: Are Voiceless Approximants Fricatives?. In *Proceedings of the 18th International Congress of Phonetic Sciences*. University of Glasgow.

Asu, E. L., Schötz, S. & Kügler, F. (2009). The acoustics of Estonian Swedish long close vowels as compared to Central Swedish and Finland Swedish. In *Proceedings of Fonetik 2009* (pp. 54-59). Department of Linguistics, Stockholm University.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Boersma, P. & Weenink, D. (2016). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.09, retrieved 1 January 2016 from http://www.praat.org/.

Bradbury, J. (2000). *Linear Predictive Coding*, Linear Predictive coding. Unpublished manuscript.

Bruce, G. (2010). *Vår fonetiska geografi*. Lund: Studentlitteratur.

Clason, K. (2009). *"Jordanese" then and now: a diachronic study of Jordanese Dutch vowels* (Master's Thesis). University of Amsterdam, Amsterdam, the Netherlands.

Dahl, Ö., Edlund, L.-E., Wastenson, L. & Elg, M. (eds.). (2010). *Sveriges nationalatlas. Språken i Sverige.* Stockholm: Norstedt

Danell, G. (1905-34). *Nuckömålet I-III*. Stockholm.

Elert, C-C. (1989). *Allmän och svensk fonetik.* Nordstedts.

Escudero, P., Boersma, P., Rauber, A., & Bion, R. (2009). A cross-dialect description of vowels: Brazilian versus European Portuguese. *Journal of the Acoustical Society of America, 126*. 1379–1393.

Flynn, N. E. J., & Foulkes, P. (2011). Comparing vowel formant normalization methods. In *Proceedings of the 17th International Congress of Phonetic Sciences.* University of Hong Kong.

Fox, R. A. & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America, 126*. 2603–2618.

Frid, J., Schötz, S. & Löfqvist, A. (2012). South Swedish diphthongisation: an articulographic and acoustic study of /u:/ in the Malmöhus dialect. In *Proceedings of Fonetik 2012*. Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg.

Gay, T. (1968). Effect of speaking rate on diphthong formant movements. *Journal of the Acoustical Society of America, 44,* 1570–1573.

Holm, G. (1980). Monoftongeringens kronologi och de finlandssvenska folkmålen. In *Studier i nordisk filologi 62* (pp. 101-107).

Hu, F. (2003). An acoustic phonetic analysis of diphthongs in Ningbo Chinese. In *Proceedings of Eurospeech 2003* (pp. 801-804). Geneva, Switzerland.

Jylhä, K. (2001). *Om hushållsordförrådet i karlebymålet*. Jyväskylä.

Kabir, A, Barker, J., & Giurgiu, M. (2010). Robust Formant Estimation: Increasing the Reliability by Comparison among three Methods. In *Proceedings of the International Conference on Circuits, Systems and Signal* (pp. 341-344). Malta.

Kim, Y. (2008). Liljendaldialektens fonologi i språkhistorisk belysning. In H. Palmén, C. Sandström and J.-O. Östman (eds.), *Dialekt i östra Nyland: Fältarbete i Liljendal med omnejd* (pp. 27-36). Nordica Helsingiensia 14. Helsinki: Nordica.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-3.

Lagman, E. (1979). *En bok om Estlands svenskar. Estlandssvenskarnas språkförhållanden.* 3A. Stockholm: Kulturföreningen Svenska Odlingens Vänner.

Lehiste, I., & Peterson, G. E. (1961). Transitions, glides and diphthongs. *Journal of the Acoustical Society of America, 38*, 268–277.

Leinonen, T. (2007). Vokaluttal i några finlandssvenska dialekter. In T. Arboe (Ed.), *Nordisk dialektologi och sociolingvistik.* (pp. 246 - 256). Århus: Aarhus University Press.

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software, 69*(1), 1-33.

Lindblad, P. (2011). *Talets akustik och perception*. Lund: Lund University.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*, 1773–1781.

Lobanov, B. M. (1971). ‘Classification of Russian vowels spoken by different speakers. *Journal of the Acoustic Society of America, 49*. 606–608.

Man, C. Y. (2007). An acoustical analysis of the vowels, diphthongs and triphthongs in Hakka Chinese. Paper presented at the International Congress of Phonetic Sciences 2007 (ICPhS XVI), August 6-10, 2007, Saarbrücken, Germany.

Maxwell, O. & Fletcher, J. (2010), The acoustic characteristics of diphthongs in Indian English. *World Englishes, 29*. 27–44.

Mayr, R. & Davies, H. (2011). A cross-dialectal acoustic study of the monophthongs and diphthongs of Welsh. *Journal of the International Phonetic Association, 41*(1), 1-25.

Miret, F. S. (1998). Some reflections on the notion of diphthong. *Papers and studies in contrastive linguistics, 34*. 27-51.

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustic Society of America, 80*. 1297–1308.

R Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Riad, T. (2014). *The Phonology of Swedish* (The Phonology of the World’s Languages). Oxford: Oxford University Press.

Roengpitya, R. (2002). Different Durations of Dipthongs in Thai: a New Finding. In *The Proceedings of the Annual Meeting of Berkeley Linguistic Society.*

Roengpitya, R. (2007) A New Look on Diphthongs in Thai. Paper Presented at the *13th Annual Meeting of the Southeast Asian Linguistics Society, SEALS XIII* (pp. 231-237). Canberra, Australia.

Schötz, S. & Asu, E. L. (2013). An acoustic study of accentuation in Estonian Swedish compounds. In Asu, E. L. & Lippus, P. (Eds.) (pp. 343-352). Peter Lang.

Schötz, S. & Asu, E. L. (2015). In Search of Word Accents in Estonian Swedish. In *Proceedings of ICPhS 2015*. University of Glasgow.

Schötz, S., Nolan, F. & Asu, E. L. (2014). An acoustic study of the Estonian Swedish lateral [ɫ]. In Heldner, M. (Ed.) *Proceedings from Fonetik 2014* (pp. 23-28). Department of Linguistics, Stockholm University.

Schulte, M. (2005). Phonological developments from Old Nordic to Early Modern Nordic I: West Scandinavian. In O. Bandle, K. Braunmüller, E. Jahr, A. Karker, H. Naumann, U. Telemann, L. Elmevik, G. Widmark (eds.). *The Nordic Languages. An International Handbook of the History of the North Germanic Languages*. Berlin, Boston: De Gruyter Mouton.

Stevens, K. N. (1998). *Acoustic Phonetics.* Cambridge, MA: MIT press.

Tasko, S. M. & Greilick, K. (2010). Acoustic and articulatory features of diphthong production: A speech clarity study. *Journal of Speech, Language, and Hearing Research, 53*, 84-99.

Thomas, E. R. & Tyler Kendall. (2007). *NORM: The vowel normalization and plotting suite.* [http://ncslaap.lib.ncsu.edu/tools/norm/]

Van Leussen, J.-W., Williams D. & Escudero P. (2011). Acoustic properties of Dutch steady-state vowels: contextual effects and a comparison with previous studies. In Lee W.-S., Zee E. (Eds.) *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1194–1197). Hong Kong.

Westerlund, R. (2010). *Ljudstrukturen i dialekten i Rödåliden: auditiv analys av fenomenen i en norrländsk dialekt i början av 2000-talet* (Licentiate Thesis) Umeå: Umeå Universitet. Institutionen för språkstudier.

Williams, D. & Escudero, P. (2014). A cross-dialectal acoustic comparison of Northern and Southern British English vowels. *Journal of the Acoustical Society of America* , 136, 2751-2761.

Yoon, T. J., Kang, Y., Han, S., Maeng, H., Lee, J., & Kim, K. (2015). A Corpus-based approach to dialectal variation in Korean vowels. In *Proceedings of the 18th Internation Congress of Phonetic Sciences.* Glasgow, UK: the University of Glasgow.

Yu, J. Li, A. & Wang, X. (2004). A contrastive investigation of diphthongs between Standard Mandarin and Shanghai accented Mandarin. *International Symposium on Tonal Aspects of Languages with Emphasis on Tone Languages* (pp. 229–234). Beijing, China.

Appendix

Table A. List of all words from which the initial dataset of 1072 diphthongs were extracted. Only diphthongs that were long, focused (foc), and primary stressed (prim) were used in the final analysis. A question mark (?) marks an unsure or unknown phonemic transcription, translation, or length category, and a hyphen (-) marks an unknown word meaning. Note that while different morphological forms of adjectives and articles are listed seperately, the English translation does not include such morphological information.

Dialect	Diphthong	Lexical item	English	Tokens	Length	Focus	Stress
Nargö	ai	/aig/	eye	5	long	foc	prim
Nargö	ai	/ain/	one, a	6	long	foc	prim
Nargö	ai	/bain/	leg	4	long	foc	prim
Nargö	ai	/bait/	bit (verb)	5	long	foc	prim
Nargö	ai	/vait/	knows	5	long	foc	prim
Nuckö	ai	?/daian/	-	1	long	foc	prim
Nuckö	ai	?/graibʉlen/	-	1	long	unfoc	prim
Nuckö	ai	/aike/	horse	1	long	foc	prim
Nuckö	ai	/aiken/	the horse	9	long	foc	prim
Nuckö	ai	/aikiar/	horses	2	long	foc	prim
Nuckö	ai	/ain/	one, a	2	long	foc	prim
Nuckö	ai	/aitː/	one, a	1	short	foc	prim
Nuckö	ai	/bainvɛrk/	leg pain	3	short	foc	prim
Nuckö	ai	/blai/	became	4	long	unfoc	prim
Nuckö	ai	/dai/	to die	3	long	foc	prim
Nuckö	ai	/daier/	dies	3	long	foc	prim
Nuckö	ai	/daiur/	dies	2	long	unfoc	prim
Nuckö	ai	/daiʉr/	dies	1	long	foc	prim
Nuckö	ai	/grait/	porridge	6	long	foc	prim
Nuckö	ai	/hai/	hay	1	long	foc	prim
Nuckö	ai	/haie/	the hay	1	long	foc	prim
Nuckö	ai	/hailasːe/	the load of hay	1	long	foc	prim
Nuckö	ai	/haim/	home	15	long	unfoc	prim
Nuckö	ai	/laiken/	the (children’s) game	1	long	foc	prim
Nuckö	ai	/lair/	mud, clay	7	long	foc	prim
Nuckö	ai	/lairjuːrd/	clay soil	1	long	foc	prim
Nuckö	ai	/maira/	more	8	long	unfoc	prim
Nuckö	ai	/rain/	clean	2	long	unfoc	prim
Nuckö	ai	/raip/	rope	1	long	foc	prim
Nuckö	ai	/sai/	to say	2	long	unfoc	prim
Nuckö	ai	/stain/	rock, stone	5	long	foc	prim
Nuckö	ai	/stainar/	rocks, stones	1	long	foc	prim

Nuckö	ai	/tufulgrait/	?mashed potatoes	1	long	foc	sec
Nuckö	ai	/vait/	knows	1	long	foc	prim
Nuckö	au	/auuna/	the eyes	2	long	foc	prim
Nuckö	au	/bauna/	bean	2	long	foc	prim
Nuckö	au	/baunana/	the beans	3	long	foc	prim
Nuckö	au	/blaut/	soft, wet	3	long	foc	prim
Nuckö	au	/blauta/	soft, wet	1	long	foc	prim
Nuckö	au	/blautː/	soft, wet	2	short	foc	prim
Nuckö	au	/graun/	-	4	long	foc	prim
Nuckö	au	/graunan/	-	1	long	foc	prim
Nuckö	au	/krausa/	lingonberries	1	long	foc	prim
Nuckö	au	/krausana/	the lingonberries	4	long	foc	prim
Nuckö	au	/krausmuːs/	lingonberry jam	1	long	foc	prim
Nuckö	au	/laus/	loose	9	long	foc	prim
Nuckö	au	/smoːgraunan/	?the blueberry	1	long	foc	sec
Nuckö	au	/smoːgraunana/	the blueberries	1	long	foc	sec
Nuckö	au	/smoːgrauns-/	blueberry-	2	long	foc	sec
Nuckö	au	/smoːgraunsmuːs/	blueberry jam	1	long	foc	sec
Nuckö	oi	/hoit/	wheat	5	long	foc	prim
Nuckö	oi	/hoit-/	wheat-	1	?	foc	prim
Nuckö	oi	/hoite/	the wheat	3	long	foc	prim
Nuckö	oi	/hoitmeːl/	wheat flour	2	short	foc	prim
Nuckö	oi	/hoitmʉːl/	wheat flour	2	short	foc	prim
Nuckö	ui	/huit/	white	6	long	foc	prim
Nuckö	ui	/huit-/	white-	1	?	foc	prim
Nuckö	ui	/huit(ː)/	white	2	?	foc	prim
Nuckö	ui	/huitː/	white	12	short	foc	prim
Nuckö	ui	/huitbreː/	white bread	1	long	foc	prim
Nuckö	ui	/huitmeːl/	wheat flour	2	short	foc	prim
Nuckö	ui	/huitʉr/	white	2	long	unfoc	prim
Nuckö	ui	/kuild/	evening	3	?	foc	prim
Nuckö	ui	/kuilde/	the evening	1	?	foc	prim
Nuckö	ui	/kuine/	the mill	6	long	foc	prim
Ormsö	ai	/aildast/	?to keep a fire	1	long	foc	prim
Ormsö	ai	/ain/	one, a	5	long	foc	prim
Ormsö	ai	/aisket/	nevertheless	1	long	foc	prim
Ormsö	ai	/aistur-/	east-	1	short	foc	prim
Ormsö	ai	/aisturbiː/	Österby	30	short	foc	prim
Ormsö	ai	/aitː/	one, a	34	short	foc	prim
Ormsö	ai	/baine/	the leg	1	long	foc	prim
Ormsö	ai	/bainvɛrk/	leg pain	35	short	foc	prim
Ormsö	ai	/bait/	bit (verb)	20	long	foc	prim
Ormsö	ai	/fail/	wrong	4	long	foc	prim

Ormsö	ai	/glaim/	to forget	4	long	foc	prim
Ormsö	ai	/grait/	porridge	2	long	foc	prim
Ormsö	ai	/gri(ː)ngrait/	?barley porridge	3	long	foc	sec
Ormsö	ai	/hail/	whole	3	long	foc	prim
Ormsö	ai	/haila/	the whole	2	long	foc	prim
Ormsö	ai	/haim/	home	1	long	unfoc	prim
Ormsö	ai	/hait/	was called	9	long	foc	prim
Ormsö	ai	/haitː/	was called	15	short	foc	prim
Ormsö	ai	/haitur/	is called	2	long	unfoc	prim
Ormsö	ai	/kurngriːnsgrait/	barley porridge	4	long	foc	sec
Ormsö	ai	/lair/	mud, clay	30	long	foc	prim
Ormsö	ai	/rainar/	-	1	long	foc	prim
Ormsö	ai	/riːsgriːnsgrait/	rice pudding	1	long	foc	sec
Ormsö	ai	/staiɫt/	-	1	short	foc	prim
Ormsö	ai	/staiɫtan/	-	1	short	foc	prim
Ormsö	ai	/vait/	knows	2	long	foc	prim
Ormsö	au	?/laudare/	-	1	long	foc	prim
Ormsö	au	/blaut/	soft, wet	28	long	foc	prim
Ormsö	au	/blautː/	soft, wet	21	short	foc	prim
Ormsö	au	/blautan/	soft, wet	1	long	foc	prim
Ormsö	au	/laudan/	the Saturday	2	long	foc	prim
Ormsö	oi	/hoit/	wheat	31	long	foc	prim
Ormsö	oi	/hoite/	wheat	1	long	foc	prim
Ormsö	oi	/hoitsmeːl/	wheat flour	22	short	foc	prim
Ormsö	oi	/hoitsmɛːl/	wheat flour	6	short	foc	prim
Ormsö	oi	/jʉːlkoilde/	Christmas Eve	2	?	foc	sec
Ormsö	oi	/koild/	evening	34	?	foc	prim
Ormsö	oi	/koilde/	the evening	3	?	foc	prim
Ormsö	oi	/koin/	mill	27	long	foc	prim
Ormsö	oi	/koine/	the mill	3	long	foc	prim
Ormsö	ui	/huit/	white	28	long	foc	prim
Ormsö	ui	/huitː/	white	34	short	foc	prim
Ormsö	ui	/huitmeːl/	wheat flour	1	short	foc	prim
Ormsö	ui	/huitsmeːl/	wheat flour	7	short	foc	prim
Rickul	ai	/aikiar/	horses	1	long	foc	prim
Rickul	ai	/ain/	one, a	6	long	unfoc	prim
Rickul	ai	/bain/	leg	2	long	unfoc	prim
Rickul	ai	/baina/	the legs	2	long	foc	prim
Rickul	ai	/baine/	the leg	3	long	foc	prim
Rickul	ai	/blai/	became	12	long	unfoc	prim
Rickul	ai	/braive/	next to	2	long	foc	prim
Rickul	ai	/dail/	part	1	long	unfoc	prim
Rickul	ai	/daiur/	dies	11	long	foc	prim

Rickul	ai	/fail/	wrong	9	long	foc	prim
Rickul	ai	/glaim/	to forget	10	long	foc	prim
Rickul	ai	/glaimd/	forgot	1	long	foc	prim
Rickul	ai	/glaimur/	forgets	1	long	foc	prim
Rickul	ai	/grait/	porridge	2	long	foc	prim
Rickul	ai	/gri(ː)ngrait/	?barley porridge	5	long	foc	sec
Rickul	ai	/gri(ː)ngraitn/	?the barley porridge	1	long	unfoc	sec
Rickul	ai	/hail/	whole	6	long	foc	prim
Rickul	ai	/haila/	the whole	4	long	unfoc	prim
Rickul	ai	/haim/	home	11	long	foc	prim
Rickul	ai	/haitː/	was called	1	short	foc	prim
Rickul	ai	/hait/	was called	1	long	foc	prim
Rickul	ai	/haitur/	is called	20	long	foc	prim
Rickul	ai	/lai/	-	1	long	foc	prim
Rickul	ai	/laid/	lead (verb)	12	long	foc	prim
Rickul	ai	/laikia(r)/	(children’s) games	1	long	foc	prim
Rickul	ai	/lair/	mud, clay	3	long	foc	prim
Rickul	ai	/mai/	May	1	long	foc	prim
Rickul	ai	/mair/	more	3	long	foc	prim
Rickul	ai	/maira/	more	6	long	foc	prim
Rickul	ai	/rainana/	-	2	long	foc	prim
Rickul	ai	/raipa/	?to pick	1	long	foc	prim
Rickul	ai	/sai/	to say	4	long	unfoc	prim
Rickul	ai	/saiur/	says	2	long	unfoc	prim
Rickul	ai	/staitur/	shoves	5	long	foc	prim
Rickul	ai	/vait/	knows	12	long	unfoc	prim
Rickul	au	?/tauan/	-	1	long	foc	prim
Rickul	au	?/sauuna/	-	1	long	foc	prim
Rickul	au	/augʉsti/	August	1	long	foc	-
Rickul	au	/bjau/	treated	1	long	foc	prim
Rickul	au	/blaut/	soft, wet	24	long	foc	prim
Rickul	au	/krausana/	the lingonberries	1	long	foc	prim
Rickul	au	/krausar/	lingonberries	10	long	unfoc	prim
Rickul	au	/krausmuːs/	lingonberry jam	3	long	foc	prim
Rickul	au	/laudan/	the Saturday	2	long	foc	prim
Rickul	au	/laup/	to walk	1	long	foc	prim
Rickul	au	/laus/	loose	21	long	foc	prim
Rickul	au	/lauser/	loose	2	long	unfoc	prim
Rickul	oi	/hoitsmeːl/	wheat flour	2	short	foc	prim
Rickul	oi	/poikar/	boys	4	long	foc	sec
Rickul	ui	/huirt/	?sharp	1	?	foc	prim
Rickul	ui	/huisːur/	sharp	4	short	foc	prim
Rickul	ui	/huitː/	white	6	short	foc	prim

Rickul	ui	/huitan/	white	1	long	foc	prim
Rickul	ui	/huitbreː/	white bread	1	?	foc	prim
Rickul	ui	/kuilde/	the evening	6	?	foc	prim
Rickul	ui	/kuin/	mill	2	long	foc	prim
Runö	ai	/baine/	the leg	4	long	foc	prim
Runö	ai	/glaime/	to forget	3	long	foc	prim
Runö	ai	/haim/	home	1	long	foc	prim
Runö	ai	/hait/	was called	6	long	foc	prim
Runö	ai	/jaim/	?home	1	long	unfoc	prim
Runö	ai	/laik/	(children’s) game	1	long	unfoc	prim
Runö	ai	/taim/	?them	3	long	unfoc	prim
Runö	ai	/vait/	knows	1	long	foc	prim
Runö	oi	/hoit/	wheat	7	long	foc	prim
Runö	oi	/hoitmeːl/	wheat flour	2	short	foc	prim
Runö	oi	/hoitsmeːl/	wheat flour	3	short	foc	prim
Runö	ui	/huit/	white	9	long	foc	prim
Runö	ui	/huit(ː)/	white	8	?	foc	prim
Rågöarna	ai	/ain/	one, a	3	long	unfoc	prim
Rågöarna	ai	/aitː/	one, a	8	short	foc	prim
Rågöarna	ai	/baine/	the leg	3	long	foc	prim
Rågöarna	ai	/bainvɛrk/	leg pain	7	short	foc	prim
Rågöarna	ai	/bait/	bit (verb)	9	long	foc	prim
Rågöarna	ai	/fail/	wrong	3	long	foc	prim
Rågöarna	ai	/haila/	the whole	3	long	foc	prim
Rågöarna	ai	/hait/	was called	5	long	foc	prim
Rågöarna	ai	/kvait/	wheat	7	long	foc	prim
Rågöarna	ai	/kvaitmjʉːl/	wheat flour	1	short	foc	prim
Rågöarna	ai	/kvaitsmjʉːl/	wheat flour	6	short	foc	prim
Rågöarna	ai	/lair/	mud, clay	6	long	foc	prim
Rågöarna	ai	/maira/	more	2	long	foc	prim
Rågöarna	ai	/rʉːnai/	Runö	3	long	foc	sec
Rågöarna	au	/austerbiː/	Österby	8	short	foc	prim
Rågöarna	au	/blaut/	soft, wet	14	long	foc	prim
Rågöarna	au	/blautː/	soft, wet	2	short	foc	prim
Rågöarna	au	/blautʉr/	soft, wet	4	long	foc	prim
Rågöarna	au	/sau/	to sow	5	long	foc	prim