

THE TIMING AND QUALITY OF DIPHTHONG COMPONENTS IN SPONTANEOUS ESTONIAN

Pärtel Lippus, Eva Liina Asu

Institute of Estonian and General Linguistics, University of Tartu, Estonia
partel.lippus@ut.ee, eva-liina.asu@ut.ee

ABSTRACT

Using data from a corpus of spontaneous speech the present study investigates the timing and quality of diphthong components in Estonian, a quantity language that combines a rich vowel system (9 monophthongs, 36 diphthongs) with complex prosodic characteristics (three quantity degrees).

In the Estonian quantity system, diphthongs are equivalent to long vowels and can occur in both long and overlong quantity degrees. Previous studies have been inconclusive as to whether the lengthening in the overlong quantity involves just one or both diphthong components.

The results of the current study, based on the formant trajectories of 9 monophthongs and 8 diphthongs from 119 speakers, show that both diphthong components are lengthened but that the lengthening is more prominent in the second component. Instead of connecting two stable targets the formant trajectories change during the course of the whole diphthong forming a constant glide.

Keywords: diphthongs, vowel quality, quantity, Estonian, spontaneous speech

1. INTRODUCTION

The aim of this paper is to analyse the timing and quality of diphthong components in spontaneous speech. More broadly, the study addresses the issue of interaction of different levels of representation – segmental with prosodic – based on Estonian, a language where these two are intricately intertwined.

In many languages the notion of a diphthong as a phonological category is relatively vague. Due to coarticulatory reasons vowels can be diphthongised. A number of studies have searched for phonetic properties that distinguish diphthongs from monophthongs, e.g. [1], [2]. Here, a diphthong is defined as a sequence of two vowels with different targets belonging to the same syllable with there being a single noticeable change in quality in that syllable [3].

The Estonian vowel inventory comprises 9 monophthongs /i, y, e, ø, æ, a, o, ɤ, u/ that can form 36 diphthongs. All 9 monophthongs can occur as

first components of a diphthong but only 5 /i, e, a, o, u/ as second components [4].

The three-way quantity system of Estonian combines durational properties of segments with tonal characteristics, e.g. [5]. Diphthongs have an important role in this system as the duration of diphthong components has been used as a basis of a prevailing quantity theory [8]. Diphthongs are treated as equivalent to long monophthongs and can thus occur in long (Q2) or overlong (Q3) quantity degrees. It has been shown in [6] that the pronunciation of the second diphthong component varies more than that of the first one. According to [7] the quality of the first component is closer to the corresponding monophthong.

The timing of diphthong components in a small variety of Estonian was studied in [8], [9]. It was shown that both components were longer in Q3 than in Q2 supporting [7] and not being in line with [10] who claimed that only the second component is longer in Q3, while the first one is not influenced by the quantity. It is implied in [6] that the durations vary depending on the diphthong but the second components are usually longer than the first ones.

The present paper addresses several research questions testing the following hypotheses:

- What is the duration of diphthongs as compared to monophthongs? Based on the phonological treatment of Estonian quantity [11] we would expect diphthongs to have the same duration as long monophthongs.
- How do the formant trajectories of diphthongs compare to those of long monophthongs? Presuming that monophthongs in Estonian are not diphthongised and their trajectories cross just one target, we would expect formant trajectories to be longer for diphthongs than for monophthongs.
- What is the timing of diphthong components in Q2 and Q3? Based on earlier research [8], [9] we would expect both components to be longer in Q3 than in Q2 diphthongs.
- What is the quality of diphthong components as compared to the corresponding monophthongs? Based on [7] we hypothesise that diphthong components are similar in their quality to the corresponding monophthongs.

Table 1: The number of tokens grouped by gender, quantity degree and vowel category.

		ɑ	æ	e	ø	ɤ	o	i	y	u	æe	ai	au	æi	ea	ei	yi	yu
F	Q2	156	108	301	35	24	259	186	23	246	21	105	91	120	20	79	21	13
	Q3	40	43	35	35	8	165	39	9	52	84	25	41	6	90	27	29	25
M	Q2	187	95	295	23	15	223	212	14	281	29	103	82	90	31	67	39	13
	Q3	57	50	45	50	9	120	55	6	63	106	32	37	6	122	24	27	27

2. MATERIAL AND METHODS

The data comes from the University of Tartu Phonetic Corpus of Estonian Spontaneous Speech [12] from 119 speakers: 62 female, 57 male (mean age 41 years, ranges from 20-85 years, sd 15).

An initial search from the corpus resulted in 15,135 disyllabic CVCV(C) words with an open first syllable. All words with creaky or breathy voice, whisper, laughter or lengthening were excluded. Only content nouns were selected for the analysis, leaving out grammatical words and verbs. Loan words and foreign names with non-initial stress were also excluded.

Using a Praat [13] script the F1 and F2 values were measured from 30 equidistant time points from the mid 60% of the duration of each vocalic segment. In order to optimise the formant analysis the method explained in [14] was used: the formant analysis was repeated with changing the formant ceiling from 4000 to 6000 Hz in steps of 10 Hz for male speakers and from 4500 to 6500 for female speakers. The settings with the lowest variance by speaker by vowel category were selected.

As in the present paper the focus is on diphthongs in Q2 and Q3, only the words in these two quantities were included.

In order to further normalise for Praat's formant analysis errors, the formant values were grouped by gender and vowel category, and the outlying values were deleted. Finally, diphthongs with fewer than 5 tokens for gender and quantity were excluded.

The final analysis is based on 5196 tokens from 9 monophthongs and 8 diphthongs. Each speaker produced a different number of tokens but on average there were 24 monophthongs and 11 diphthongs per speaker. The number of tokens grouped by gender, quantity degree and vowel category is given in Table 1.

Statistical analysis was carried out in R using lme4 package [15]. Each acoustic measure was tested with a linear mixed effects model for the effects of quantity and vowel type (monophthong vs. diphthong), including random intercepts for the speaker and the vowel.

3. RESULTS AND DISCUSSION

Figure 1 presents the mean formant trajectories grouped by speaker gender, quantity and vowel category in F1-F2 space. Despite the different area of the vowel space of male and female speakers due to the differences in the length of the vocal tract, the distribution of the vowels is generally very similar.

3.1. The duration of diphthongs

The duration of the whole vocalic segment (see Figure 2) was 122 ms in Q2 and 155 ms in Q3 [$F(1, 4979.5) = 597.9$; $p < 0.001$]. The difference between monophthongs and diphthongs was not significant. This confirms our hypothesis and lends experimental support to the treatment of Estonian diphthongs as comparable to long monophthongs.

3.2. Formant trajectory length (TL)

Formant trajectory length (TL) was measured in two ways. Firstly, we calculated the sum of the Euclidean distances between the consecutive measuring points within a vowel. There was a significant difference between monophthongs and diphthongs [$F(1, 15.2) = 23.3$, $p < 0.001$] and a main effect of quantity [$F(1, 5133.0) = 62.6$, $p < 0.001$]. TL was 1.6 barks in Q2 vs. 1.8 barks in Q3 for monophthongs and 2.6 barks in Q2 vs. 2.8 barks in Q3 for diphthongs.

Secondly, we calculated the Euclidean distance between the first and the last measuring point within each vowel. As the first method also captures the quivering of the trajectory around the target value, the second method provides a more clear-cut distinction between monophthongs and diphthongs [$F(1, 15.0) = 37.1$, $p < 0.001$]. The distance between the start and end of the trajectory in monophthongs was 0.8 barks (quantity not significant) and in diphthongs 1.9 barks in Q2 and 2.1 barks in Q3 [$F(1, 5145.6) = 42.2$, $p < 0.001$].

As the formant trajectories were longer in diphthongs than in monophthongs, it can be concluded that monophthongs in Estonian are not diphthongised.

Figure 1: Vowel formant trajectories in F1-F2 space on bark scale from female (top panels) and male speakers (bottom panels) extracted from words in Q2 (left) and Q3 (right). Grey points show the mean formant values of monophthongs. Red points denote the starting point of a diphthong, connected with the rainbow-coloured trajectory to the blue dots showing the end point of a diphthong.

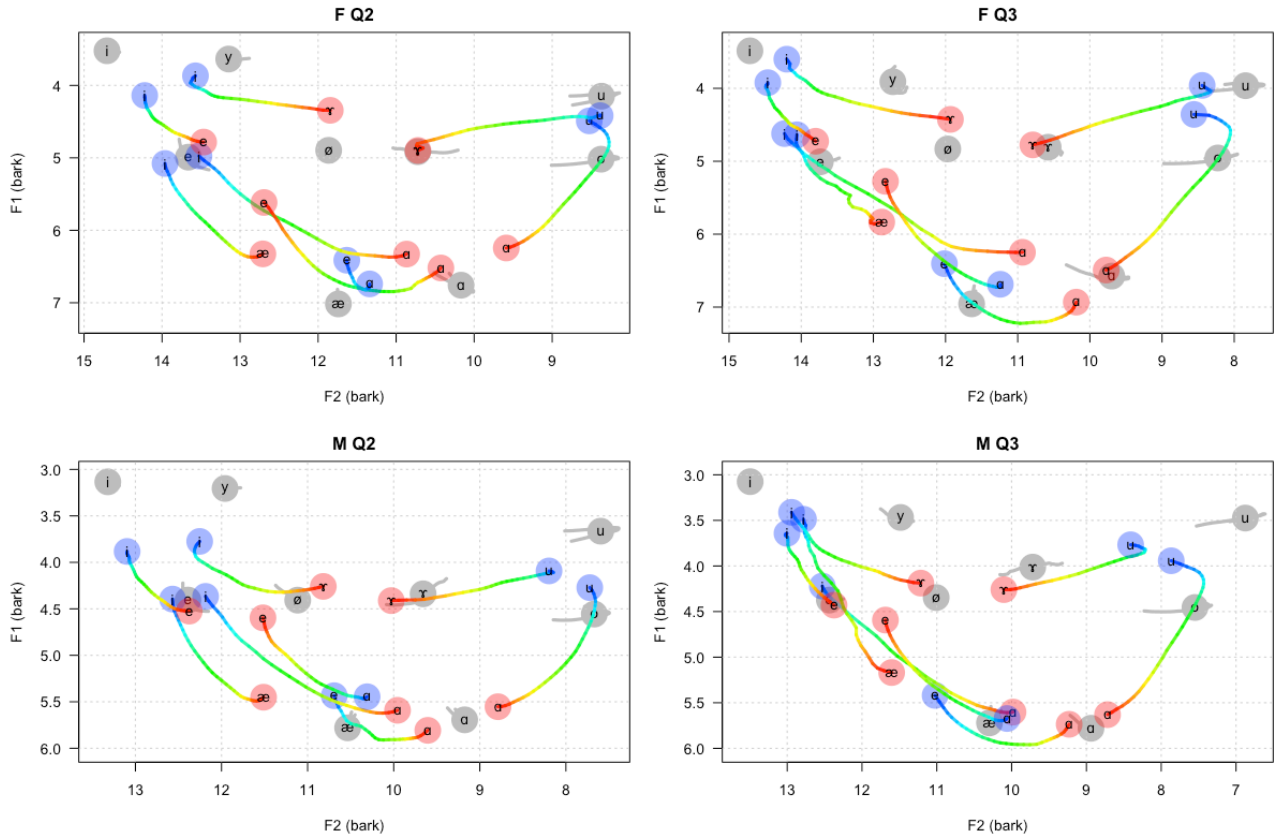
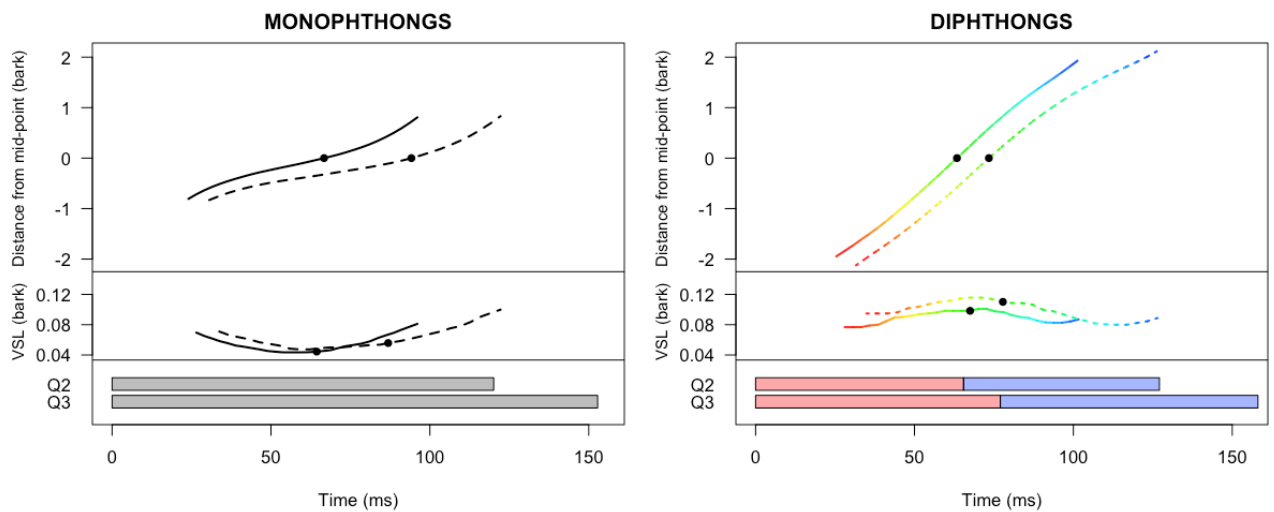


Figure 2: The temporal properties of monophthongs (left) and diphthongs (right). The upper sections show distance from the mid-point of the trajectories which is marked with a dot. The middle sections show the Vowel Section Length (VSL) with the average maximum marked with a dot. The bottom sections show the whole duration of the vowels and for diphthongs the manually annotated boundary between the first (red) and the second (blue) component. Solid line denotes Q2, dashed line Q3.



The formant trajectories of monophthongs circle around the target value of the vowel while the trajectories of diphthongs have different start and end points implying that there is a clear distinction between monophthongs and diphthongs in Estonian.

3.3. Timing of diphthong components

Based on the manual segmentation of the corpus, the duration of the first component was 65 ms in Q2 and 77 ms in Q3, the quantity effect being significant [F(1, 1571.3) = 94.6, $p < 0.001$]. The duration of the second component was 62 ms in Q2 and 81 ms in Q3 [F(1, 873.03) = 250.6, $p < 0.001$]. Thus, proportionally the manually labelled boundary was at 51% of the whole vowel duration in Q2 and at 49% in Q3 (with a significant quantity difference at [F(1, 1582.8) = 43.6, $p < 0.001$]).

Two alternative measures were used to compare with the manually annotated boundary. Firstly, Vowel Section Length (VSL), which is the rate of change (in bark) of the formant trajectory, is displayed in the middle section of Figure 2. If a diphthong consists of two stable sections at target values connected by a transition, the peak of the transition should be at the point where VSL reaches its maximum. In our data the maximum point of VSL was highly variable. The mean values were very similar to the means of the manually detected boundary, but the distribution of the points was uniform. Thus, it seems that the formant trajectories consist of a constant glide instead of two stable targets connected with a transition.

Secondly, the boundary between the diphthong components as a mid-point in the formant trajectory is displayed in the top section of Figure 2. This was calculated as the point which is at an equal distance from the beginning and end of each vowel (Euclidean distance in bark values). The mid-point of the formant trajectory was at 51% of the vowel duration in Q2 and at 48% in Q3 [F(1, 1621.2) = 10.9, $p < 0.001$].

The results show that both diphthong components were affected by the quantity. As expected, the diphthong components were longer in Q3 than in Q2 confirming the results of a study based on read materials from a small Estonian dialect [8], [9]. The lengthening in the second component was, however, slightly more prominent. The results were also affected by vowel intrinsic properties.

3.4. Target undershoot

Figure 3 shows the Euclidean distance of each diphthong component from the target values, i.e. the corresponding monophthongs.

Figure 3: The distance of the first (left panel) and the second (right panel) diphthong component from the target vowel in bark. Highlighting shows the expected closest candidate (purple), mismatch between the expected and the closest candidate (blue), and an alternative closer target (red).

	a	æ	e	ø	ɤ	o	i	y	u		a	æ	e	ø	ɤ	o	i	y	u
ae	0.9	1.4	3.7	2.5	2	2.4	5.2	3.9	3.1	ae	2	0.9	2.1	1.4	1.9	3.7	3.7	2.7	4.1
ai	1.2	1.2	3	1.9	1.7	2.8	4.5	3.3	3.4	ai	3.8	2.7	0.9	1.7	2.9	5	1.8	1.5	5.2
au	0.8	2	4.1	2.7	1.9	1.8	5.6	4.1	2.5	au	2.5	3.7	4.9	3.4	2.2	1.1	6	4.4	1.1
æi	2.7	1.4	1.6	1.6	2.5	4.4	3.3	2.5	4.8	æi	4.1	2.9	0.9	2	3.2	5.4	1.7	1.6	5.6
ea	2.9	1.9	1	1	2.1	4.3	2.5	1.6	4.5	ea	1.3	0.9	2.8	1.8	1.8	3.1	4.4	3.2	3.6
ei	3.8	2.7	0.6	1.6	2.8	5	1.8	1.4	5.2	ei	4.7	3.6	1.2	2.4	3.6	5.8	0.9	1.5	5.9
ɤi	2.8	2.3	1.7	1	1.6	3.5	2.8	1.5	3.6	ɤi	4.4	3.5	1.3	2	3.2	5.3	1	1.2	5.3
ɤu	2	2.1	2.6	1.2	0.9	2.5	3.9	2.4	2.7	ɤu	2.7	3.6	4.7	3.2	2	1.3	5.7	4.1	1.1

In most cases the first component of the diphthong is closest to its target vowel and average distance from the target value is one bark. However, in few cases the diphthong component is close to several targets: in [ai] the first component is between [a] and [æ], and in [ea] it is between [e] and [ø]. In the case of [ɤi] the first component is closer to [ø] than to its target [ɤ].

In the case of the second component we can see more evidence of undershoot. The average distance of the second component from the target vowels is 1.4 barks. In the case of [ai] and [æi] the second target is lowered and is closer to [e] than to [i] (0.9 vs. 1.7-1.8 barks). In the case of [æe] and [ea] the second component is closest to [æ] (0.9 barks in both cases vs. 2.1 and 1.3 barks respectively).

In line with [6] it can be seen that the target values of diphthong components do not always coincide with the corresponding monophthongs.

4. CONCLUSIONS

The present study investigated several aspects of diphthongs in Estonian. It was shown that there is a good match between the phonological category and phonetic realisation. Rather than connecting two steady targets, the formant trajectories change continually throughout the whole diphthong.

Regarding the temporal properties, both diphthong components lengthen in Q3, but the second component is more affected by the quantity. However, more systematic study is needed in order to further analyse the effects of intrinsic properties on the timing of diphthong components.

5. ACKNOWLEDGEMENTS

This research was supported by the National Program for the Estonian Language Technology project EKTB3. The paper is dedicated to the 101st anniversary of the Republic of Estonia.

6. REFERENCES

- [1] L. Feng, 'The phonetic boundary between monophthongs and diphthongs in Suzhou Chinese', *Toronto Working Papers in Linguistics*, vol. 28, pp. 49–56, 2008.
- [2] E. Jacewicz, R. A. Fox, and J. Salmons, 'Vowel change across three age groups of speakers in three regional varieties of American English', *Journal of Phonetics*, vol. 39, no. 4, pp. 683–693, 2011.
- [3] I. Lehiste and G. E. Peterson, 'Transitions, Glides, and Diphthongs', *J. Acoust. Soc. Am.*, vol. 33, no. 3, pp. 268–277, 1961.
- [4] E. L. Asu and P. Teras, 'Illustrations of the IPA. Estonian', *Journal of the International Phonetic Association*, vol. 39, no. 03, pp. 367–372, Nov. 2009.
- [5] P. Lippus, E. L. Asu, P. Teras, and T. Tuisk, 'Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian', *Journal of Phonetics*, vol. 41, no. 1, pp. 17–28, Jan. 2013.
- [6] H. Piir, 'Acoustics of the Estonian diphthongs', *Estonian Papers in Phonetics*, vol. 1982–1983, pp. 5–103, 1985.
- [7] I. Lehiste, 'Diphthongs versus vowel sequences in Estonian', in *Proceedings of the Sixth International Congress of Phonetic Sciences, September 7-13, 1967, Prague*, B. Hála, M. Romportl, and P. Janota, Eds. Prague: Academica, 1970, pp. 539–544.
- [8] E. L. Asu, P. Lippus, E. Niit, and H. Türk, 'The acoustic characteristics of monophthongs and diphthongs in the Kihnu -variety of Estonian', *Linguistica Uralica*, vol. 48, no. 3, pp. 161–170, 2012.
- [9] H. Türk, E. L. Asu, P. Lippus, and E. Niit, 'Diftongidest ja triftongidest eesti keeles Kihnu vokaalisüsteemi näitel', *Keel ja Kirjandus*, vol. 59, no. 1, pp. 1–12, Jan. 2016.
- [10] A. Eek, *Eesti keele foneetika*. Tallinn: Tallinna Tehnikaülikooli Kirjastus, 2008.
- [11] T.-R. Viitso, 'Phonology, morphology and word formation', in *Estonian language*, M. Ereht, Ed. Tallinn: Estonian Academy Publishers, 2003, pp. 9–92.
- [12] P. Lippus, P. Teras, T. Tuisk, and N. Ots, 'Phonetic Corpus of Estonian Spontaneous Speech v.1.0.4'. Center of Estonian Language Resources, 14-Feb-2018. Retrieved from <https://dx.doi.org/10.15155/1-00-0000-0000-0000-00154L>
- [13] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. (Version 6.0.43). 2018. Retrieved from <http://www.praat.org>
- [14] P. Escudero, P. Boersma, A. S. Rauber, and R. A. H. Bion, 'A cross-dialect acoustic description of vowels: Brazilian and European Portuguese', *The Journal of the Acoustical Society of America*, vol. 126, no. 3, pp. 1379–1393, Sep. 2009.
- [15] D. Bates, M. Mächler, B. Bolker, and S. Walker, 'Fitting Linear Mixed-Effects Models Using lme4', *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.