The intonation of Saaremaa Estonian: in search of Swedish influence

Eva Liina Asu
Department of Linguistics, University of Cambridge, UK

Abstract
The variety of Estonian normally singled out as differing from the standard in its intonation is that of the island of Saaremaa (Ösel) in the Baltic Sea. Its characteristic melody has been attributed to direct contacts with Swedish and claimed phonetically to be manifested in a higher post-stressed syllable as compared to the stressed syllable. In the present paper, a direct comparison is made on the basis of read speech between the intonation of Saaremaa and Standard Estonian. Contrary to the hypothesis of peak delay in the Saaremaa variety no significant difference is found in peak alignment between the varieties. The data show, however, that there is a significant difference in the height of initial pitch both at the beginning of an utterance and at the beginning of the second accent unit.

Introduction
The first systematic description of Estonian intonation (Asu, 2004) concentrates solely on the analysis of the standard variety. The present paper is the first step in investigating regional variation in the intonation of Estonian. Of immediate interest when tackling intonational variation in Estonian is the intonation of the varieties spoken on the islands off the west coast of the Estonian mainland. According to Lonn and Niit (2002: 11) the speech melody on every major Estonian island differs slightly. The variety normally singled out is that of the island of Saaremaa (Ösel), which in the literature is frequently described impressionistically as “sing-songy”. At the same time surprisingly little phonetic research has been carried out in order to pinpoint the phonetic characteristics of this “sing-songiness”. A sociolinguistic study by Koreinik (2001) explains this phenomenon as being due to the later rise of the F0 contour but does not give a reference to any experimental studies on the matter.

Niit (1980), in her study of the word prosody of Estonian coastal and island varieties, has shown that in these varieties the unstressed syllable following the stressed syllable is higher in pitch. She suggests that the characteristic melody of the Saaremaa variety is due to direct contacts with Swedish in the Baltic Sea prosodic area, or Sprachbund, where language contacts between genetically unrelated languages have been noticed and studied since Jakobson (1931) (e.g. Lehiste 1988, Wiik 1997).

Niit’s work enables Niit and Remmel (1985) to hypothesize that Estonian coastal and island dialects form a transitional area between Standard Estonian and Swedish with regard to tone application: the F0 contour (presumably of a disyllabic word) in these varieties is an intermediate form between Swedish accent 1 (early peak as in Standard Estonian), and accent 2 (generally characterized by a late peak).

In sum, the preceding work implies that in the island varieties of Estonian the F0 peak is delayed, and that there is a later rise in the F0 contour. However, as far as is known no direct comparative studies have been carried out. The present paper sets out to fill this gap and compare the intonation of Standard Estonian and Saaremaa with the specific aim of testing the hypothesis about peak alignment. F0 peaks in Saaremaa are expected to be aligned later than in the standard reflecting the possible Swedish influence on the prosody of the island variety.

Methodology

Materials
The data used comprise read declarative sentences that were controlled for the quantity opposition and the number of accents. Each utterance consists of four accent units. The present study focuses on the comparison of peak alignment in the first two H*+L accent units in the utterances. The amount of material used for the study of the two accent units differs. The data-set for the comparison of the initial accent is larger consisting of repetitions of four different utterances where the first accent unit comprises a disyllabic (Q2) proper noun (e.g. Leena or Miina) followed by a disyllabic verb in an un-
stressed position (e.g. lamab ‘is lying’ or valab ‘is pouring’). For the comparison of the second accent the repetitions of only one of the four utterances were chosen where the second accent unit consists of a tetrasyllabic word Jaamusega (‘with Jaanus’). Not all of the tokens could be used because, as shown in Asu (2002), the default declarative intonation pattern of a sequence of H*+L accents can often change within an intonation phrase yielding low accentuation at different points after the first two accents and influencing the realization of the preceding H* accent.

Subjects and procedure

From each variety six speakers were recorded. They were all female between the ages of 21 and 35. The recordings of Standard Estonian were made in Tartu where all subjects were university students. Saaremaa recordings were conducted in Kuressaare where only those subjects were recorded who were born and had lived in the region for most of their life. All recordings were made in a quiet environment using a Sony TCD D8 portable DAT recorder and a dynamic microphone with a cardioid response. The subjects were asked to read the utterances from a list where each utterance appeared five times in a semi-randomized order.

The data were digitized at 16 kHz using Xwaves+™ running on a Silicon Graphics Unix workstation. An F0 contour and a wideband spectrogram were computed for each utterance.

For the initial accent unit each F0 contour was quantified in terms of eight measurement points at which also the time value was recorded: (1) the initial pitch, (2) the beginning of the stressed vowel, (3) the mid-point of the stressed vowel, (4) the end of the stressed vowel, (5) the peak i.e. the highest value in the F0 contour, (6) the mid-point in the unstressed vowel in the proper noun, (7) the mid-point in the first unstressed vowel of the verb, (8) the mid-point in the second unstressed vowel of the verb. Eight comparable measurements were also taken in order to quantify the F0 contour of the second accent unit. The first measurement point there was taken at the onset of the /j/ sound at the beginning of the word Jaamusega (‘with Jaanus’).

In order to minimize differences in pitch span for different speakers with higher and lower voices the measurements in Hz were converted into semitones (re 50 Hz).

Results

Peak alignment

Utterance initial peak

Figure 1 presents the average contours of utterance initial peaks for each variety. The duration of the F0 contour in the figure is normalized relative to the length of the accent unit (which in this case is expressed as a stretch from the beginning of the F0 contour to the mid vocalic measurement of the last unstressed vowel in the accent unit). The solid line represents the mean values for all six speakers of the standard variety and the dashed line the mean values for all six speakers of the Saaremaa variety.

It can be seen in Figure 1 that in both varieties the peak is delayed until after the end of the stressed vowel (marked by a shaded area). The peak in the Saaremaa variety is aligned slightly later than that in the standard variety.

Figure 1. Average F0 contours (means of all speakers) of the utterance initial peaks for the two varieties: Standard Estonian (solid line) and Saaremaa (dashed line). The shaded area marks the stressed vowel.

A further comparison of the peak alignment of individual speakers reveals that the patterns for the two varieties are indeed very similar. In the data of most of the speakers the initial peak is delayed until after the stressed vowel.

In order to reliably compare the data of different speakers and to deal with speaker-dependent differences in speaking rate the location of the peak was normalized and expressed as a percentage of the accent unit. The results of this normalization are shown in Figure 2. It can be seen that there is considerable variation within in the data. A paired-samples t-test carried out on the means indicates no significant
The overall patterning of the data in Figure 3 is similar to that of the first peak in Figure 2. The alignment of the peaks of the Saaremaa speakers seems to be on the whole slightly later in the accent unit than that of the Standard Estonian speakers. However, a paired samples t-test fails to give a significant result between the two varieties (t(5) = 1.048; p>0.05).

**Pitch height**

More striking than the slight difference in peak alignment between the two averaged contours in Figure 1 is the difference in the height of initial pitch. In order to compare this difference in a reliable way the initial pitch height was expressed as a percentage of the span of an accent unit in relation to the F0 peak and the final measurement in the accent unit. This was done separately for each accent unit.

Figure 4 and Figure 5 show the results of normalized initial pitch for the two accent units. On the whole, for both accent units, the initial pitch is higher for the speakers of Standard Estonian than than for the Saaremaa speakers.

Paired samples t-tests show that there is a significant difference between the two varieties with respect to the height of accent unit initial pitch (t(5) = -2.709; p<0.05) for the first accent unit.
Discussion and conclusions

The hypothesis about peak delay based on earlier research was not borne out in the present data as no significant difference was found between the normalized peak alignment of the two varieties. This applies to both the initial peak and the second peak. The non-significant result could be influenced by at least two factors that are to do with the choice of speakers and the data used for the comparisons.

According to Koreinik (2001) the characteristic “sing-songy” intonation is evident in the speech of only about 50% of the inhabitants of Saaremaa who have lived on the island for at least three generations. This figure is based on purely subjective evaluation of the interviewees’ speech but implies that a tighter control is needed of the subjects’ background. In order to capture the “pure” regional accent it might be useful to carry out an informal evaluation of the subjects’ speech before recording.

Another factor that is likely to have influenced the result of the peak alignment is the formal speaking task used for the present experiment. It is well known that regional variation becomes more evident in spontaneous speech than in read speech, and therefore the highly controlled utterances might not be the best test bed for differences in intonational variation.

However, the present data show a significant difference in the height of initial pitch at the beginning of both of the accent units analyzed. On average Saaremaa speakers start the F0 contour lower than speakers of Standard Estonian. The result is somewhat unexpected as there is no reference to this in previous literature. On the other hand the finding links to some previous research on intonational variation that has shown the importance of the characterization of the overall shape of the pitch accent gesture (Aufterbeck, 2003).

It is planned in further research to investigate the phonetic realization of the intonation of Saaremaa on the basis of spontaneous speech, which might additionally reveal phonological differences. In order to expand on the sociolinguistic dimension of the study it is also planned to include both male and female speakers of at least two age groups, as well as speakers from different parts of the island.

In the longer term the aim of the work undertaken is to carry out a comparison with Swedish in order to test the hypothesis about the influence of Swedish on the prosody of Estonian islands.

Acknowledgements

I am very grateful to Viive Torn for her help in recruiting the subjects in Kuressaare, Saaremaa. I would also like to thank Francis Nolan for discussing various aspects of the research presented in this paper.

References


