THE EFFECT OF INTONATION ON PITCH CUES TO THE ESTONIAN QUANTITY CONTRAST

Eva Liina Asu* and Francis Nolan‡
*Department of Estonian, University of Tartu, Estonia
‡Department of Linguistics, University of Cambridge, UK

ABSTRACT

Estonian has a three-way quantity contrast operating at the level of the foot, e.g. [sada] ‘hundred’, [saada] ‘send (2 sg. imperative)’, and [saa:da] ‘get (infinitive)’. It has been known for some time that, in citation forms at least, pitch as well as duration is correlated with the contrast. In particular the ‘overlong’ third quantity is associated with an early pitch peak, and a concomitant fall in pitch on the first syllable. This finding, however, has apparently been made on the basis of prepausal tokens spoken with falling intonation. The experiment reported here investigates whether there is a parallel pitch cue when the intonation is rising. This turns out not to be the case, and the paper discusses the reasons why this might be, and what the theoretical implications are.

1. INTRODUCTION

Estonian is well known as a language which has a three-way quantity contrast, traditionally referred to as short (Quantity 1), long (Quantity 2), and overlong (Quantity 3), demonstrable in minimally distinct triplets such as Q1 *sada* [sada] ‘hundred’, Q2 *saada* [saada] ‘send (2 sg. imperative)’, and Q3 *saada* [saa:da] ‘get (infinitive)’. The analysis of quantity has been much discussed (e.g. [5, 12]). Eek (e.g. [3]) has argued that the domain of the contrast is the disyllabic foot, and that what is crucial perceptually is the durational ratio between the stressed syllable nucleus (plus coda, if any) and the nucleus of the immediately following unstressed syllable. According to traditional grammar monosyllabic words are regarded as being of Q3.

Duration, however, is not the only phonetic dimension associated with the quantity contrast. There are many references (e.g. [9, 11, 3]) to a supporting difference in the pitch contour. In particular it has been noted that Q3 is associated with a marked fall in pitch during the first syllable, or an early placement of a pitch peak, whereas the other two quantities have a step down between the two syllables [10:243]. Krull [8], however, has shown that this difference is robust only in prepausal words in sentence final position.

Commenting on a variety of work related to the quantity contrast Gårding [4:409] says

> It is obvious that much more attention is devoted to rhythm (quantity and stress) than to intonation, which is only mentioned in terms of Fo contours described by the location of Fo peaks in the syllable of one fixed (probably declarative) intonation pattern. Would a study of the influence of intonation on stress and quantity patterns only add to the confusion? I think it might help in establishing the constant properties by which the contrastive quantity patterns are recognized.

The experiment reported here is a first step towards exploring the interaction of utterance-level intonation with the quantity system. It compares the pitch contours of isolated words on falling and rising intonation contours. Rising intonation contours are not frequent in Estonian, perhaps because ‘yes-no’ questions, a common context for rises in some languages, are in effect ‘wh-’ questions, being marked at the start by the word *kas* ‘whether’. Nevertheless rises do occur, commonly for instance on the utterance *mida?* ‘what/pardon?’, and they are natural (though not inevitable) on non-final items in lists. It is therefore reasonable to suppose that speakers are able to realise rises as part of their intonational repertoire.

2. EXPERIMENT

The materials consisted of sequences of segmentally matched words contrasted by quantity such as *kalu* [kalu] ‘fish (part. pl.)’, *kaalu* [kaalu] ‘weight (gen. sing.)’, *kaalu* [kaa:lu] ‘weight (part. sing.)’. The distinction between Q2 and Q3, which is not represented in the orthography, was indicated by a superscript ‘II’ or ‘III’; since the quantity distinctions are dealt with explicitly in school, this is adequate to elicit the right form from educated Estonian speakers. All words used were of the form CV[l]V, on the grounds that the medial resonant [l] would minimise Fo perturbations. Five such sets of segmentally matched words were used: *koli, tuli, kalu, kilu, vala* and their Q2 and Q3 counterparts. Each set was read six times, exemplifying each possible ordering of the three quantities, as shown in the table below. The order in which these ‘lines’ were read was also varied from word to word.

I	II	III
I	III	II
II	I	III
II	III	I
III	I	II
III	II	I

Subjects tended naturally to read the first two tokens in a line as rising and the last one as falling. Where they were inconsistent, or misread a line in some way, they were asked to re-record the line.

The subjects were three educated female speakers of standard Estonian aged between 25 and 28. Recordings were made in a quiet environment on a Marantz PMD221 analogue cassette recorder.

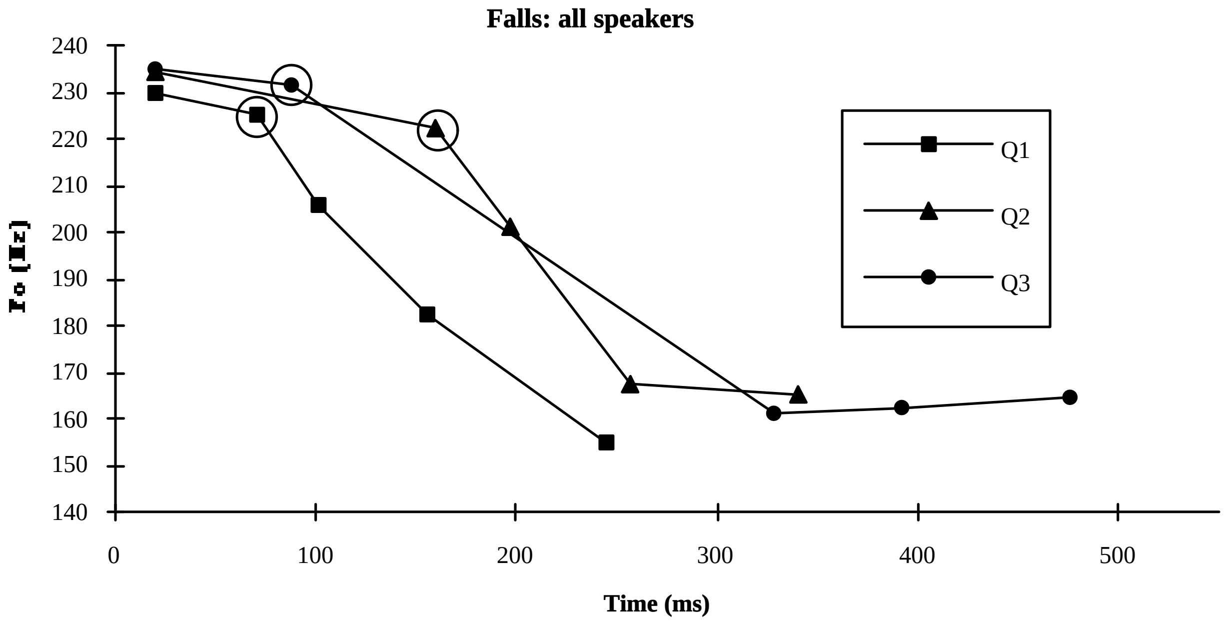

Figure 1. Mean fundamental frequency values at the five measurement points for each quality: falling intonation, all words, three speakers. The circled point is the turning point.

3. ANALYSIS

The recordings were digitised at 16 kHz using Xwaves+ running on a Silicon Graphics Unix workstation in the phonetics laboratory of Cambridge University Linguistics Department. A fundamental frequency contour and a spectrogram were computed for each token. From each line of data only the middle token (manifesting a non-terminal rise) and the final token (manifesting a final fall) were analysed. In this way, as can be seen in the above table, there were two rising tokens and two falling tokens available for each segmentally matched sequence on each quantity (e.g. [koo:li]). Thus, in the data presented below, a value for 'rising Q3' is the average of 2 tokens x 5 words x 3 speakers = 30 observations. Although it can be argued that the fall in each group of three tokens is 'utterance final' and the rise 'utterance medial', and so not directly comparable, both are at least phrase-final since subjects read the words with a major prosodic boundary (marked by a pause) between each item.

Preliminary observation of the Fo contours, waveforms, and spectrograms suggested that the falling contours might be adequately characterised by Fo at five definable events: the start of the first vowel, the 'turning point', the start of the lateral, the start of the second vowel, and the end of the second vowel. The time of these events was also recorded. In measuring Fo some discretion was used to avoid unrepresentative Fo values caused by consonantal perturbations, either after the word-initial plosive or adjacent to the lateral. In cases where the value at the measurement point was out of line with the trend of adjacent points, a more representative value was estimated from the adjacent values.

The definition of the 'turning point' requires some explanation. For falling contours, it represents either a change in direction from rising to falling, or from level to falling, or, more contentiously, a change in trajectory from a shallow to a steep fall. Where the word exhibited a fall at a constant rate from the beginning of the word, the turning point coincides with the beginning of the vowel. For rising contours, the turning point represents the start of a steeply rising contour after either a level or a shallow rise. Again, in the case of a consistent rise where no change in steepness could be observed, the turning point would coincide with the beginning of the word. Turning points were most problematic to determine on the relatively short contours of Q1, and the high variance of these values suggests that they should be treated with more caution than those for the longer contours.

4. RESULTS

Figure 1 plots the five measured Fo values for each of the three quantities. The values are plotted on a real-time axis, with the value for the start of the first vowel plotted at 20 ms (chosen arbitrarily to reflect the short aspiration of the initial stop). Data for the three speakers is averaged in Figure 1.

It can be seen, most strikingly, that whereas for Q1 and Q2 the turning point (the second point plotted) is aligned half or two-thirds of the way through the first vowel (i.e. towards the start of the lateral, the third point plotted), in Q3 the turning point is early in the long vowel, at similar absolute duration to that of the Q1 turning point. As a result there is a considerable (and clearly audible) fall in pitch during the first vowel of Q3 which is not evident for Q1 and Q2, where the pitch falls only gradually within the first vowel, the pattern being more of a step down between syllables – both findings corresponding to previous descriptions such as those cited in the Introduction.

As far as duration is concerned, Figure 1 shows more or less expected relationships. For Q1 the durations of the first vowel and the second are 82 and 89 ms (yielding a V1:V2 ratio of 0.9); for Q2 177 and 83 ms (2.1); and for Q3 308 and 84 ms (3.7). These ratios follow the same trend as previously

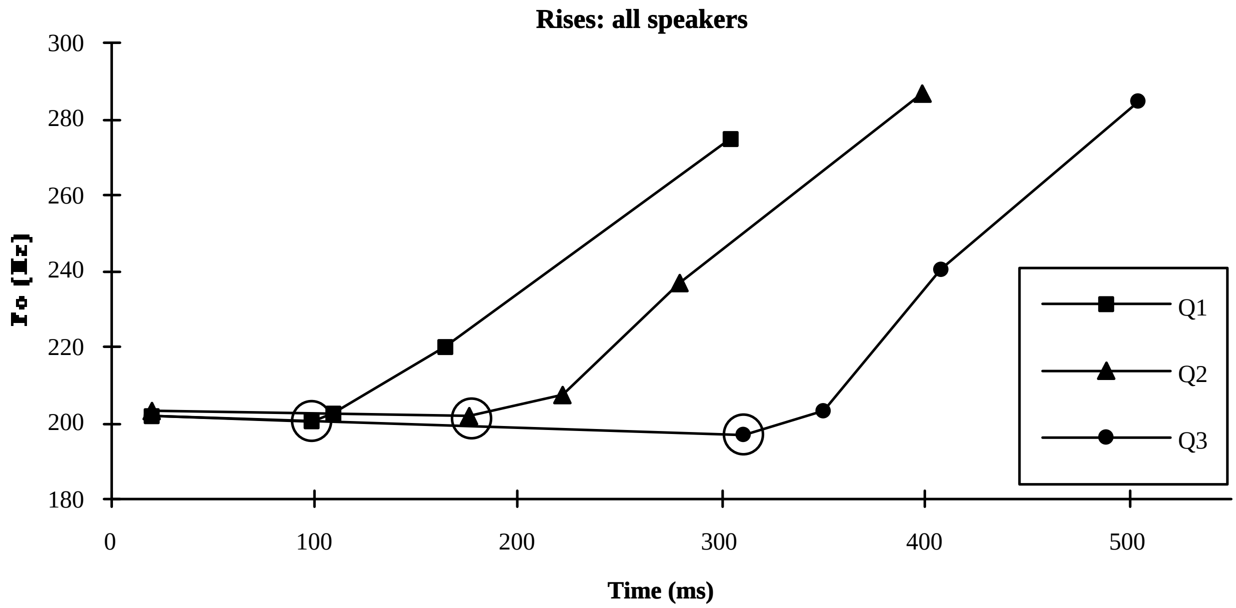

Figure 2. Mean fundamental frequency values at the five measurement points for each quality: rising intonation, all words, three speakers. The circled point is the turning point.

observed, as shown in Table 1 taken from [3] with the ratios of the present study added. Some degree of variation in the ratios across different studies is not surprising, given for instance the conditioning effects on durations of the variety of segmental structures which the studies will have used.

	Q1	Q2	Q3
Asu & Nolan	0.9	2.1	3.7
Lehiste [9]	0.7	1.5	2.0
Liiv [13]	0.7	1.6	2.6
Eek [2]	0.7	2.0	3.9
Krull [6, 7]	0.5–0.7	1.2–2.1	2.2–2.9

Table 1. Duration ratios associated with the three quantities found in a number of studies (after [3])

The main question is whether the early turning point of Q3 is an alignment specific to the fall, or arises from a constant principle of intonational realisation which would be reflected in the alignment of an equivalent (but mirror image) turning point in the rising contours. This bears on the general question of the alignment of intonational targets, dealt with for instance in [1].

Figure 2 answers this question in a straightforward way. It can be seen that the alignment of the turning point is, in all quantities, very close to the end of the first vowel. It seems as if the first syllable has to be low, and only near the onset of the lateral of the second syllable can the pitch start to rise – which it continues to do through the second vowel.

The duration ratios for rises are compared with those for falls in Table 2.

	FALLS			RISES		
	V1	V2	V1:V2	V1	V2	V1:V2
Q1	82	89	0.9	89	140	0.6
Q2	177	83	2.1	202	118	1.7
Q3	308	84	3.7	330	97	3.4

Table 2. Durations of the first and second vowels (ms), and the approximate ratios between them

It can be seen that there is general agreement in the trend. The first syllable should be, for Q2, around twice the duration of the second syllable, and in Q3, around 3.5 times the duration of the second syllable. In the case of Q1 the second syllable is the longer, though more dramatically so in the rise. Overall, the second vowel tends to be longer relative to the first when the intonation is rising.

To facilitate the statistical analysis of the turning points of rises and falls each turning point was expressed as a proportion of the duration of the V1 in which it occurred. The mean of these values is given in Table 3.

	FALLS	RISES
Q1	56.2	92.2
Q2	79.1	80.1
Q3	21.6	88.1

Table 3. Time of mean turning point expressed as a proportion of the duration of the first vowel

A repeated measures Analysis of Variance with the independent variables *intonation* (falling, rising), *quantity* (1,3) and *subject* (1,3), and the dependent variable *turning point* was carried out. The analysis revealed significant main effects of intonation ($F[1,9]=72.05$, $p<0.001$), quantity ($F[2,18]=15.09$, $p<0.001$) and subject ($F[2,18]=8.02$, $p<0.01$), and significant interactions between all three independent variables. Of primary interest to this paper is the interaction between intonation and quantity ($F[2,18]=24.9$, $p<0.001$) because it suggests that quantity affects turning point in a different way when intonation rises or falls. Planned comparisons showed that within the falling pitch movements, turning points differed significantly from each other in the three quantities (Q1,Q2 $p<0.01$, Q1,Q3 and Q2,Q3 $p<0.0001$), but within the rising pitch movements, the turning points did not differ significantly in the three quantities. As noted above the difficulty of measuring the turning point in Q1 in the falls may mean that less weight should be put on values involving Q1; but the Q2-Q3 difference is highly reliable, and represents clear support for a distinct placement of the turning point in falls.

There is, conversely, no evidence for a quantity-correlated difference of principle in the alignment of pitch targets for the rise. The turning point appears to be anchored near the onset of the second syllable, even in Q3.

5. DISCUSSION AND CONCLUSIONS

Broadly it can be concluded that the durational cues to the three quantities are robust in the face of intonational variation. On the other hand the previously observed pitch cue to the third quantity, involving an early turning point and a steep fall on the first syllable, has no analogue in the rising contours. The start of the steep rise seems to be aligned near the end of the first vowel whatever the quantity of the word, and, apart from the trivial fact that the duration of low pitch material differs with the length of the vowel, pitch has no part to play in cuing the different quantities when the intonation is rising. Why should there be this asymmetry between rises and falls?

Lehiste [10:243] tentatively explains the distinctive Q3 falling pattern as follows:

> In many instances, overlength...is the result of the loss of a vowel after a long syllable... Disyllabic sequences without overlength on the first syllable have a step-down pitch contour distributed over the two syllables. It may well be that the loss of the vowel of the second syllable resulted in transferring the tonal contour of the whole disyllabic sequence to the lengthened first syllable, which thus acquired its distinctive falling pattern.

In autosegmental terms, perhaps, the low tone associated with the original second syllable survived the loss of the vowel and reassociated to the first syllable. Perceptual reinterpretation would then have led to the early fall becoming conventionalised as part of the package of cues to the third quantity.

Could lack of the pitch correlate under rising intonation be regarded as a case of phonological neutralisation? Clearly not, since the three-way phonological distinction continues to be carried by duration when a rising intonation is used. But neither can the lack be seen as comparable to the contextually determined absence or presence of a secondary cue in certain segmental contrasts. The fortis-lenis contrast in English stops is arguably cued principally by the aspiration of the voiceless stops, since initial and final /b d g/ tend not to have vocal cord vibration; but in intervocalic position voicing acts as an additional cue. This distribution of cues, while not inevitable given that many languages have fully voiced initial stops and no aspiration, does follow fairly naturally from the aerodynamics and physiology of the vocal mechanism.

The distribution of the secondary pitch cue to the quantities, on the other hand, is more clearly under the direct control of the speaker. There is no *phonetic* reason why the turning point should be earlier in the falling third quantity. It is not, for instance, preserving a minimum distance from the end of the utterance, to ‘allow room’ for the fall, since it places the turning point much further from the end of the utterance in Q3 than in Q1 and Q2. Nor is there any phonetic obstacle to an early turning point in the rising intonation Q3. The status of the early turning point, in terms of synchronic phonological representation, remains a puzzle.

In historical terms it is perhaps not surprising that a pitch cue derived from the loss of a syllable should be retained in the case of an intonation pattern which is heard and produced frequently, but not in a relatively rare intonation pattern. Linguistic irregularities, for instance irregular or suppletive verb morphology, are much more likely to survive in common items. In the case of the rarer (rising) intonation pattern speakers globally apply the pitch alignment which seems to be the general pattern for Estonian, with the exception of falling Q3, namely that the first syllable has one tone and the second syllable the other (H L or L H).

ACKNOWLEDGMENT

We are grateful to Geoff Potter and Esther Grabe for advice and assistance in the preparation of this paper.

REFERENCES

[1] Arvaniti, A., Ladd, D.R. and Mennen, I. 1998. Stability of tonal alignment: the case of Greek pre-nuclear accents. *Journal of Phonetics* 26, 3-25.

[2] Eek, A. 1974. Observations on the duration of some word structures: I. *Estonian Papers in Phonetics, EPP 1974* 18-31.

[3] Eek, A. and Meister, E. 1998. Estonian speech in the Babel multilanguage database: phonetic-phonological problems revealed in the text corpus. *Proc. of the Workshop on Speech Database Development for Central and Eastern European Languages*, Granada, May 1998.

[4] Gårding, E. 1998. Review of ‘Estonian Prosody: Papers from a Symposium’. *Journal of Phonetics* 26, 401-409.

[5] Hint, M. 1997. *Typological Problems of Estonian Grade Alternation and Prosodical System.* Tallinn/Helsinki: Eesti keele sihtasutus.

[6] Krull, D. 1991. Stability in some Estonian duration relations. *PERILUS* 13, 57-60. Institute of Linguistics, University of Stockholm.

[7] Krull, D. 1992. Temporal and tonal correlates to quantity in Estonian. *PERILUS* 15, 17-36. Institute of Linguistics, University of Stockholm.

[8] Krull, D. 1997. Prepausal lengthening in Estonian: evidence from conversational speech. In Lehiste, I. and Ross, J. (eds), Estonian Prosody: Papers from a Symposium. Tallinn: Institute of Estonian Language, 136-148.

[9] Lehiste, I. 1960. Segmental and syllabic quantity in Estonian. *American Studies in Uralic Linguistics* 1, 21-82. Indiana University, Bloomington.

[10] Lehiste, I. 1978. Polytonicity in the area surrounding the Baltic Sea. In Gårding, E., Bruce, G., and Bannert, R. (eds), *Nordic Prosody: Papers from a Symposium*. Department of Linguistics, Lund University.

[11] Lehiste, I. 1997. The structure of trisyllabic words. In Lehiste, I. and Ross, J. (eds), Estonian Prosody: Papers from a Symposium. Tallinn: Institute of Estonian Language, 149-164.

[12] Lehiste, I. and Ross, J. (eds) 1997. *Estonian Prosody: Papers from a Symposium.* Tallinn: Institute of Estonian Language.

[13] Liiv, G. 1961. Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid. *Keel ja Kirjandus* 7, 412-424, 8, 480-490.