Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

F0 declination in spontaneous Estonian: implications for pitch-related preplanning in speech production

Eva Liina Asu [1], *Pärtel Lippus* [1], *Nele Salveste* [1], *Heete Sahkai* [2]

[1] Institute of Estonian and General Linguistics, University of Tartu, Estonia
[2] Institute of the Estonian Language, Estonia

eva-liina.asu@ut.ee, partel.lippus@ut.ee, nele.salveste@ut.ee, heete.sahkai@eki.ee

Abstract

This study contributes to the discussion on pitch-related preplanning in spontaneous speech production. It investigates the relationship of phrasal length with declination slope, and the initial and final F0 height in intonation phrases extracted from a corpus of Estonian dialogues. The analysis is based on data from 10 speakers. The results show that the declination in shorter phrases is steeper than in longer ones, and that the phrase-initial F0 is higher in longer phrases. The phrase-final F0 height was, however, found to be lower in longer phrases implying that the pitch range in longer phrases is wider. The findings of the study are discussed with reference to the nature of spontaneous speech and preplanning in speech production more generally.

Index Terms: declination, F0, preplanning, intonation phrases, spontaneous speech, Estonian

1. Introduction

The phenomenon whereby pitch gradually descends during an intonation phrase (IP) has been observed across a great number of languages, and is usually referred to as F0 declination [1]. On the one hand, F0 declination has been connected to the articulatory mechanisms behind speech production being a by-product of speech breathing ([2], [3]), but on the other hand, it has been linked to linguistic preplanning in speech [4] and is believed to signal prosodic coherence within intonation phrases [5].

Declination can be modelled as linear regression lines fitted to local F0 peaks and valleys of the F0 contours, called the top line and baseline ([1], [6]), or to all points in the F0 contour, as the overall regression line or midline (e.g. [7], [8]). The top line interacts with phonological categories controlled by the laryngeal muscles, whereas the baseline reflects the global component of an intonation phrase controlled by subglottal pressure ([9], [10]). For example, Yuan and Liberman [6] show for Mandarin Chinese and English that the two languages differ in top line and baseline trends in that the lines are similar in English but follow slightly different trends in Chinese. For Estonian it was shown [11] that the top line varies as a function of focus position: if focus is located at the beginning of the sentence, the top line is strongly negative, if at the end of the sentence, the top line is less steep or there is no declination at all. Top line and baseline slopes can show different trends depending on the phonological properties of a language (e.g. pitch accents or lexical tone are more likely to interfere with the top line than with the baseline).

The issue of phrasal preplanning of F0 is controversial as different studies have yielded varying results, some supporting global (phrasal) preplanning (or 'hard' preplanning) but others local preplanning. Phrasal preplanning of F0 declination (or the lack of it) is evidenced in the relationship between utterance duration and declination slope, as well as utterance duration and the height of the initial F0 peak. For some languages longer IPs have been shown to correlate with less steep declination slope (e.g. [8] for Swedish). It has, however, been argued in [4] that the correlation between utterance duration and (the top line) declination does not necessarily support the existence of phrasal preplanning, and that the peaks in a downstepping contour can be quantified by means of a downstep ratio, i.e. a constant fraction of the preceding peak, the difference in slope being the result of the consequent asymptotic lowering of pitch.

Studies on the relationship between utterance duration and the height of the initial F0 peak have not provided unanimous support for the existence of phrasal preplanning. Initial F0 peaks have been shown to be higher in longer utterances (e.g. [6], [12], [13]), or not to depend on utterance duration (e.g. [4]). Prieto et al. [13] found large speaker-dependent variation in their data of five Romance languages, which they interpret as an argument against global preplanning and evidence for 'soft' preplanning, i.e. preplanning as an optional choice by the speaker (see [4] for the term).

Preplanning of F0 declination has also been shown to depend on utterance type. For instance, it has been demonstrated for read Danish [14] that statements have the steepest and unmarked questions the shallowest slope. Prenuclear slope was also found to be significantly different in various utterance types in Estonian [15], where the slope in yes/no-questions was similar to that of statements, but was less steep in tag-questions and unmarked questions, providing evidence for the trade-off relationship between intonation and morphosyntax. The same study showed on the basis of rigorously controlled read data that F0 downtrends in Estonian statements revealed no time-dependent effect on the scaling of F0 peaks (in line with [4]).

On the whole, more evidence has been found for preplanning in read speech as compared to spontaneous speech, where declination is more variable and less frequently present (e.g. [5], [7], [8], [16]). This is explained by the fact that spontaneous speech is planned simultaneously with speaking (the so called 'plan-as-you-go' model) and there is less opportunity for a look-ahead [5].

As the majority of studies on declination have dealt with controlled read speech, the need for investigating declination in large corpora of spontaneous speech in order to shed light to

 doi: 10.21437/SpeechProsody.2016-234

pitch-related preplanning in spontaneous speech has been pointed out (e.g. [6]).

The aim of the present study is to describe F0 declination in spontaneous conversational Estonian by focusing on the relationship between utterance duration and declination slope, and utterance duration and phrase-initial F0 height. We assume that higher phrase-initial F0 and a less steep declination slope in longer IPs would give evidence of an intention to maintain declination throughout the IP, e.g. as a signal of its internal coherence. Hence this would support the idea that preplanning is involved even in spontaneous speech.

2. Materials and method

2.1. Speech data

The present investigation is based on data extracted from the Phonetic Corpus of Estonian Spontaneous Speech (http://www.keel.ut.ee/foneetikakorpus/). The corpus contains about 60 h of spontaneous speech. For the purposes of the analysis speech data from 5 dialogues was used including 10 speakers (5 female and 5 male). The speakers were between 21 and 32 years old (average age 25 years) and represent the variety of Standard Estonian. About 5 minutes of speech was analysed for each speaker.

2.2. Method

In the current working model for the prosodic labelling of Estonian, the existence of a major intonation phrase boundary can be signalled by the following features: a pause, segmental lengthening, creaky voice and a clearly perceivable pitch reset. As the corpus of spontaneous speech is segmentally annotated (including information about segmental lengthening and voice quality) a Praat script was used to automatically locate the prosodic phrase boundaries based on the above-listed first three boundary criteria. Automatic prosodic labelling was manually checked by four independent transcribers in particular with the view of pitch reset as a boundary feature. All questionable cases were solved by reaching a consensus among the transcribers.

Declination slopes were calculated automatically by the contour-based parametric superpositional F0 analysis of local pitch events associated with the stressed syllable [17], [18]. The F0 contours were interpolated, stylised (see [17], [18] for details), and converted to semitones (st) based on the formula:

$$F0st = 12\ log2(F0Hz/b) \quad (1)$$

where the base value (b) is the F0 median below the 5th percentile of an utterance.

Three linear regression lines were fitted to the F0 contours: top line, midline, and baseline. The F0 declination was calculated by collecting data in the frames of 200 ms shifted along the F0 vector with the step size of 10 ms. Within each window, the following values were extracted: F0 median based on the values below the 10th percentile for the baseline, F0 median based on the values above the 90th percentile for the top line, and F0 median based on all the values. The intercept of the regression describes the pitch range from the phrase onset to the local minimum of the phrase, and the slope shows the rate of declination.

In addition to the declination slopes phrase-initial maximum and phase-final minimum F0 values were calculated as the 95th percentile within the initial 25% of the phrase, and as the 5th percentile within the final 25% of the phrase. These values were then normalised to the semitone scale using the formula given in (1) with reference to the speaker's overall mean F0 value (b).

3. Results

In total there were 945 IPs in the data. The midline slope was negative in 716 (75.8%) IPs, while the top line and baseline slopes were negative in 724 (76.6%) and in 666 (70.5%) IPs respectively. The final analysis was carried out only with the phrases where all three slopes were negative, i.e. all in all 584 phrases. The duration of the analysed IPs ranged from 0.2 to 5.4 sec, with the median duration 1.2 sec.

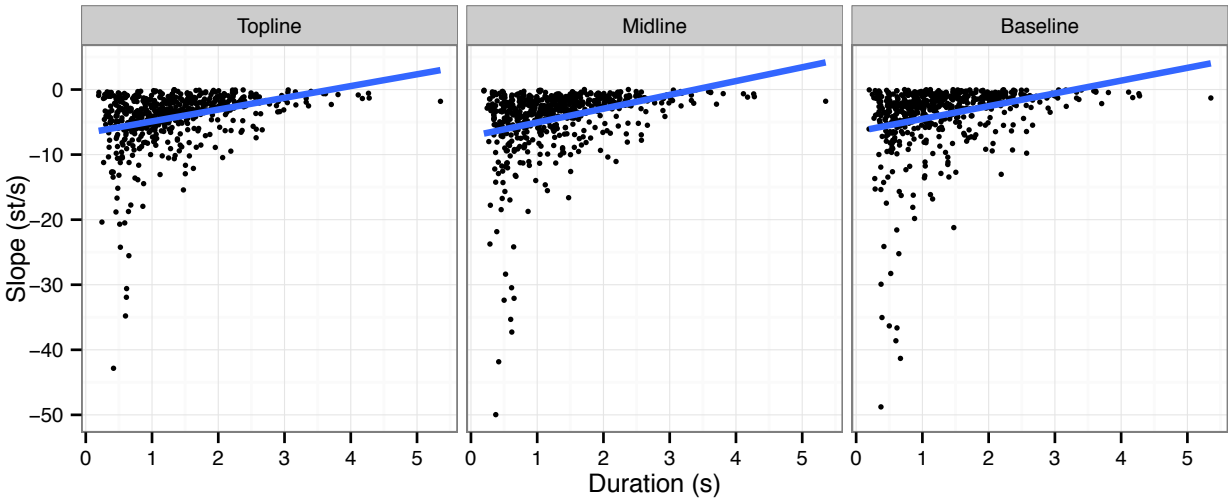

Figure 1: *Top line, midline and baseline slopes (in semitones per second) as a function of IP duration.*

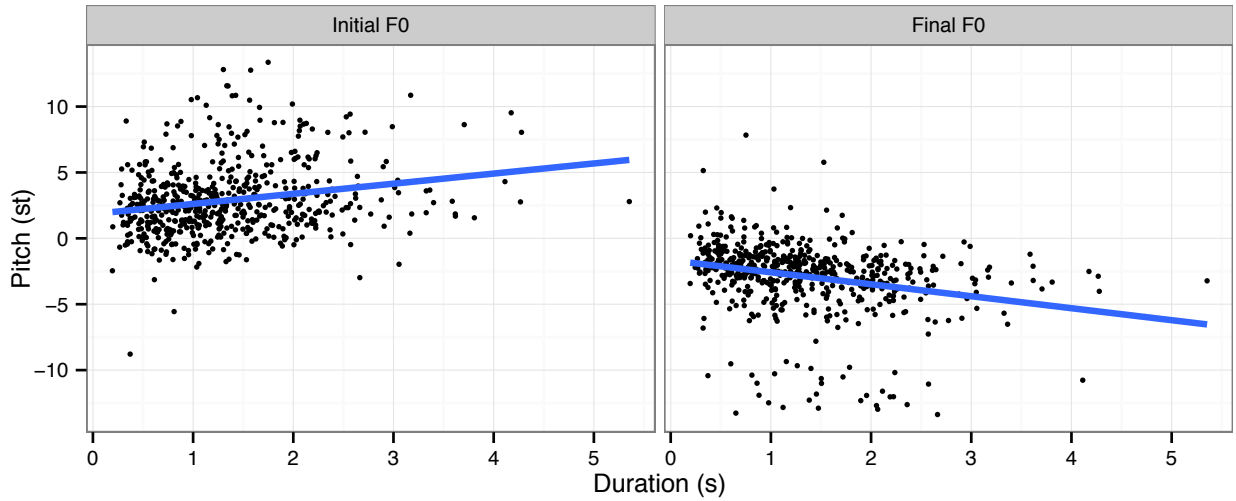

Figure 2: *Phrase-initial and phrase-final F0 height (in semitones) as a function of IP duration.*

3.1. Declination slope

The mean intercept for the top line was 9.6 st, for the midline 8 st, and for the baseline 6 st. Mixed effects linear models showed no significant effect of phrase duration on the intercept: top line [χ^2 (df = 4, N = 584) = 2.973; p = 0.08], midline [χ^2 (df = 4, N = 584) = 0.278; p = 0.6] and baseline [χ^2 (df = 4, N = 584) = 0.023; p = 0.9].

Mixed effects linear regression models showed, however, a significant effect of duration on the slope: top line [χ^2 (df = 4, N = 584) = 91.896; p<0.001], midline [χ^2 (df = 4, N = 584) = 102.53; p<0.001] and baseline [χ^2 (df = 4, N = 584) = 87.408; p<0.001]. Figure 1 illustrates the three slopes as a function of phrase duration. It can be seen that the downtrends in all three slopes are very similar. On average, the declination is about 4 st/sec (-4.3 st/sec for the top line, -4.4 st/sec for the midline, and -3.9 st/sec for the baseline). There is a general trend in all three slopes for the initially negative slope to become less steep as the phrase duration increases. The top line slope loses its steepness by 2.1 st/sec as the phrase lengthens by one second, the midline slope by 2.5 st/sec, and the baseline slope by 2.3 st/sec.

3.2. Phrase-initial and phrase-final F0 height

Figure 2 presents the phrase-initial and phrase-final F0 (with reference to the speaker's mean F0) as a function of time. There is a significant effect of phrase duration on both the phrase-initial F0 [χ^2 (df = 4, N = 584) = 25.625; p<0.001] and the phrase-final F0 [χ^2 (df = 4, N = 583) = 33.304; p<0.001]. The initial F0 height increases with increasing phrase duration by about 0.7 st/sec, while at the same time the phrase-final F0 decreases by 0.7 st/sec.

This data show that longer intonation phrases start at a higher pitch level than shorter ones. Interestingly, phrase-final F0 lowered as a function of phrase length. The raising of the initial F0 and lowering of the final F0 with increasing phrase duration points to greater pitch range in longer utterances.

4. Discussion

The present study revealed a correlation between phrasal length and declination slope: the declination slope (as expressed by three linear regression lines: top line, midline and baseline) was less steep in longer intonation phrases than in shorter ones. This could be interpreted as an effect of phrasal preplanning, but could just as well be the inevitable consequence of the so called local declination model [4], where the F0 of each accent is a proportion of the previous one giving an asymptotic decline.

The three different slopes studied exhibited almost identical rates of declination. In this sense, the Estonian data are comparable to a recent study on English where the top line and baseline were shown to be similar [6]. The fact that the top line did not show more variation than the baseline also indicates that in the present data, effects of focal pitch accents (as reported in [11]) did not interfere with the overall downtrend. This could be explained by the general nature of the data and the methodologies used.

Several studies have found that the relationship between phrasal duration and declination, as well as duration and phrase-initial F0 is weaker in spontaneous than in read speech (e.g. [8]). In a number of studies on read speech, initial F0 peaks have been shown to be higher in longer utterances, which has been treated as a strong evidence for pitch-related preplanning (e.g. [6], [12], [13]). The results presented here demonstrate that the phrase-initial F0 height depends on the duration of the phrase, implying thus that even in spontaneous speech the speakers are well aware of the length of the upcoming utterance, and plan their F0 accordingly. The present finding is therefore not in line with e.g. [5], where no correlation between F0 starting point and phrasal length was shown for spontaneous Swedish. However, further study is needed in order to determine the level of speaker variation in the data. A high degree of variation between speakers was found within several Romance languages by [13] who took it as evidence for the so called soft preplanning, concluding that

tonal preplanning is an optional mechanism (like pitch range choices) in speech production.

It was shown in the present study that the final F0 varies together with the duration of the intonation phrase, being lower in longer phrases than in shorter ones. This result is somewhat surprising, because most previous studies have reported the final low to be constant for the speaker [4] and for the phrases of different length [6]. These studies are, however, all based on read or partly scripted speech. The present study investigated spontaneous conversational speech, which is inherently much more interactive. In spontaneous dialogues, a speaker is often interrupted or s/he gives just short feedback to the interlocutor. Therefore, shorter intonation phrases are more likely to end at higher pitch within the speaker pitch range, whereas longer phrases, being more probably uninterrupted finished utterances, end at the speaker's low pitch. The present results showing a higher phrase-final F0 in shorter phrases can thus easily reflect the interactive character of spontaneous conversations used for the analysis, considering that the number of shorter intonation phrases was much larger in the dataset than the number of longer phrases.

The issue of preplanning in relation to utterance type was not explored in the present study but remains of interest in future studies. The data controlled for utterance type is expected to additionally show different downtrends for different utterance types (as in [15]). Also, a comparison of read and spontaneous Estonian data using similar methodology would help to clarify further the controversial issue of preplanning in read and spontaneous speech.

5. Conclusions

This study investigated the relationship between utterance duration and declination slope, and utterance duration and phrase-initial F0 height in spontaneous Estonian. The analysis showed that the declination slope (as expressed by three linear regression lines: top line, midline and baseline) was shallower in longer intonation phrases as compared to shorter phrases. Likewise, phrase-initial F0 height was shown to depend on the duration of the phrase: the longer the phrase the higher the phrase-initial F0. These results can be interpreted as evidence for the presence of phrasal preplanning in spontaneous speech. It was additionally shown that the phrase-final F0 varied together with the duration of the intonation phrase being higher in shorter utterances, and implying that the speakers use a wider pitch range in longer phrases.

6. Acknowledgements

We are grateful to Uwe Reichel for his generous help with the data analysis. This work was supported by Estonian Research Council grants IUT2-37 and IUT35-1.

7. References

[1] A. Cohen, R. Collier, and J. 't Hart, "Declination: Construct or intrinsic feature of speech pitch?," *Phonetica*, vol. 39, no. 4–5, pp. 254–273, 1982.

[2] R. Collier, "Physiological correlates of intonation patterns," *The Journal of the Acoustical Society of America*, vol. 58, no. 1, pp. 249–255, 1975.

[3] S. Maeda, "A characterization of American English intonation," Ph.D. thesis, Massachusetts Institute of Technology, 1976.

[4] M. Liberman and J. Pierrehumbert, "Intonational invariance under changes in pitch range and length," in *Language Sound Structure*, M. Aronoff and R. T. Oehrle, Eds. Cambridge, MA: MIT Press, 1984, pp. 157–233.

[5] P. Hansson, *Prosodic phrasing in spontaneous Swedish*. Lund: Lund University, 2003.

[6] J. Yuan and M. Liberman, "F0 declination in English and Mandarin Broadcast News Speech," *Speech Communication*, vol. 65, pp. 67–74, Nov. 2014.

[7] P. Lieberman, W. Katz, A. Jongman, R. Zimmerman, and M. Miller, "Measures of the sentence intonation of read and spontaneous speech in American English," *The Journal of the Acoustical Society of America*, vol. 77, no. 2, pp. 649–657, 1985.

[8] M. Swerts, E. Strangert, and M. Heldner, "F0 declination in read-aloud and spontaneous speech," in *Proceedings of the fourth International Conference on Spoken Language Processing, USA, Philadelphia*, 1996, vol. 3, pp. 1501–1504.

[9] J. 't Hart, R. Collier, and A. Cohen, *A perceptual study of intonation an experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press, 1990.

[10] K. Honda, "Physiological factors causing tonal characteristics of speech: from global to local prosody," in *Speech Prosody*, Nara, Japan, 2004.

[11] N. Salveste, "On the pragmatic and semantic functions of Estonian sentence prosody," Ph.D. thesis, Ludwig-Maximilian-Universität München, 2015.

[12] A. Rialland, "Anticipatory raising in downstep realization: Evidence for preplanning in tone production," in *Proceedings of the Symposium Cross-Linguistic Studies of Tonal Phenomena. Tonogenesis, Japanese Accentology, and Other Topics*, S. Kaji, Ed. Tokyo: Tokyo University of Foreign Studies, 2001, pp. 301–321.

[13] P. Prieto, M. D'Imperio, G. Elordieta, S. Frota, and M. Vigario, "Evidence for soft preplanning in tonal production: initial scaling in Romance," in *Proceedings of the 3rd International Conference Speech Prosody*, R. Hoffmann and H. Mixdorff, Eds. Dresden: TUDpress, 2006, pp. 803–806.

[14] N. G. Thorsen, "A study of the perception of sentence intonation—Evidence from Danish," *The Journal of the Acoustical Society of America*, vol. 67, no. 3, pp. 1014–1030, 1980.

[15] E. L. Asu, "The phonetics and phonology of Estonian intonation," Ph.D. thesis, University of Cambridge, 2004.

[16] J. Tøndering, "Preplanning of intonation in spontaneous versus read aloud speech: evidence from Danish," in *Proceedings of the 17th International Congress of Phonetic Sciences*, W.-S. Lee and E. Zee, Eds. Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong, 2011, pp. 2010–2013.

[17] U. Reichel, "The copasul intonation model," in *Elektronische Sprachverarbeitung*, B. Kroeger and P. Birkholz, Eds. Dresden: TUDpress, 2011, pp. 341–348.

[18] U. D. Reichel and K. Mády, "Comparing parameterizations of pitch register and its discontinuities at prosodic boundaries for Hungarian," in *Proceedings of Intercpeech*, Singapore, 2014.