

F_0 DECLINATION IN READ-ALOUD AND SPONTANEOUS SPEECH

Marc Swerts[†], Eva Strangert[‡] and Mattias Heldner[‡]

[†]Institute for Perception Research (IPO), Eindhoven, The Netherlands

[‡]Department of Phonetics, Umeå University, Sweden

ABSTRACT

This paper deals with a prosodic comparison of spontaneous and read-aloud speech. More specifically, the study reports data on F_0 declination in these two speaking modes using Swedish materials. For both speaking styles the analysis revealed negative slopes, a steepness-duration dependency with declination being less steep in longer utterances than in shorter ones and resetting at utterance boundaries. However, there was a difference in degree of declination between the two speaking styles, read-aloud speech in general having steeper slopes, a more apparent time-dependency and stronger resetting than spontaneous speech.

1. INTRODUCTION

There is a growing interest in comparative studies of read-aloud and spontaneous speech. These are generally thought to represent two ‘speaking styles’ that are so distinct that results obtained for one do not necessarily generalize to the other. The study to be reported on here aims at testing whether the two styles are prosodically different. More specifically, it concentrates on the analysis of various aspects of declination, defined as the tendency for fundamental frequency (F_0) to gradually drift downward in the course of an utterance. The aspects investigated include (i) size of declination slope, (ii) steepness-duration dependencies comparing differently-sized units of speech and (iii) extent of boundary resetting.

Declination is particularly interesting to investigate from a comparative perspective, because it has been claimed to be very much style-dependent ([4], [10], [11], [16]). On the one hand, declination has been shown to function as a structuring device, since a reset in the slope of declination often coincides with boundaries of various linguistic units ([12], [15]). However, such results are generally based on analyses of isolated, read-aloud sentences; some researchers think it is questionable that they are also true of spontaneous types of speech. In such a view, declination is criticized “as being purely an artefact of the specially wooden style in which informants tend to read sentences when asked to do so in decontextualised experimental conditions” ([4], p. 127).

Before embarking on the actual research, a few words need to be said about the specific choices we made for this study. First, we realize that a simple dichotomy between read and spontaneous speech is probably too rough a distinction for research purposes. Indeed, within the category of spontaneous speech, one might, for instance, distinguish between simple small-talk and a well-prepared lecture, two speaking styles probably having significantly different prosodic patterns. Alternatively, it may be more useful to classify speaking styles according to a dimension of preparedness. Following [1], the read materials discussed below probably reflect speech prepared beforehand and the spontaneous a less prepared style of speaking.

Second, apart from the fact that declination is argued to be style-dependent, it is also controversial, because it is difficult to reliably measure it. Ideally, one would like to have a declination line that can be interpreted in linguistic terms, e.g. as a reference line to which local pitch movements can be scaled. Most earlier studies, however, have relied on eye-fitting procedures, the results of which can easily be biased by the researcher’s goals; also, such results are intrinsically difficult to verify objectively. Recently, more quantitative measures have been introduced ([8], [11]), all based on some form of regression analysis. But they in turn have been criticized ([5], [9]), partly because it is unclear whether the obtained declination fits have some psychological and/or perceptual relevance. Given these disclaimers, we nevertheless opt to use an automatic and quantitative estimation procedure, because such a method is reproducible.

Finally, we need to address the question about the ‘domain’ of declination. In the literature, it has been reported that declining F_0 trends may occur over linguistic units of various sizes, from lower-level phrases ([6], [12]) up to larger-scale discourse segments ([15]). For Swedish, declination in terms of a successive stepwise downdrift at the accents is included as a characteristic feature of intonation models for utterances (e.g. [2]). Downdrift of F_0 peaks and valleys also characterize whole text units ([3]). In the present study, we have therefore chosen to concentrate on experimentally determined units that differ in size, i.e., phrases and utterances.

2. METHOD

2.1. Materials

We analyzed two short Swedish speech samples produced by a male speaker: a read-aloud news telegram (233 words) and a spontaneous retelling of the same text (252 words). The same materials were used in a previous labelling study [14] in which nine transcribers scored the strength of boundaries between successive words on a 4-point scale (0: no boundary; 3: paragraph boundary). Based on mean scores of the transcribers, two kinds of units were isolated, one defined as a stretch of speech ending with a mean boundary of at least 2 (utterance), and the other with a mean boundary of at least 1 (phrase).

Below, fragments are given of the read and spontaneous monologue, provided with the experimentally obtained phrase and utterance boundaries, visualized with single and double slashes respectively:

- (*read-aloud*)
“enligt libyska uppgifter föll / åtta fyrahundrafemtiokilosbomber / över Tripoli och Bengazi / när de amerikanska bombplanen slog till natten till tisdagen // en av bomberna föll bara ett tiotal meter / från den byggnad där ledaren Muammar Gaddafis familj låg och sov / i en militärförläggning nära Tripoli // en av dem som dödades var Gaddafis sexton månader gamla adoptivdotter Hanna / som bodde med familjen i Tripoli // ...”
- (*spontaneous*)
“det amerikanska luftangreppet sattes in i tisdags // och det genomfördes främst utav bombare som startade från England // under själva anfallet / så stördes det libyska luftförsvaret ut elektroniskt utav amerikanska sändare på fartyg i Medelhavet utanför Tripoli // så att själva luftangreppet blev ju då en framgång / libyerna kunde inte försvara sig // mot Tripoli fällde man åtta stycken fyrahundrafemtiokilosbomber // ...”

Note that earlier analyses of similar speech materials have shown that the difference in boundary strength is reflected in pause length (silent interval duration) as well as in the minimum F_0 value before a pause, being lower the higher the rank of the boundary [13]. Additional support for silent interval duration and also for final lengthening as parameters of boundary strength in Swedish is to be found in [7].

2.2. Measurements

As already said in the introduction, there are as yet no uncriticized methods available to quantitatively determine the slope and the domain of declination. Automatic procedures proposed so far all appear to be based on linear regression techniques. Lieberman *et al.* [11] calculated three different declination measures, by fitting linear regression lines to local

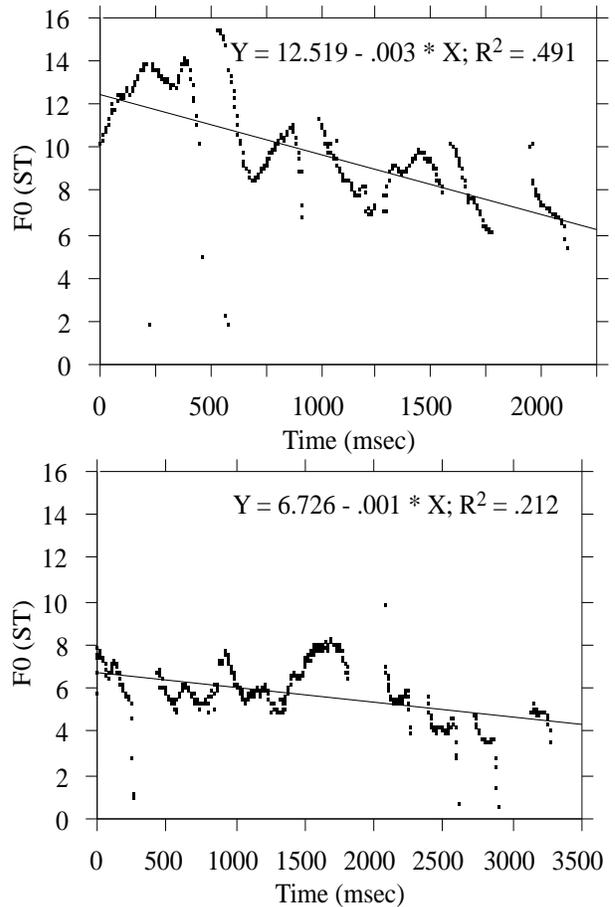


Figure 1: Examples of regression line fits to estimate the rate of declination in a read-aloud (top) and spontaneous (bottom) phrase.

peaks (topline), local valleys (baseline) and all F_0 points. They argue that the all-points line is a better descriptor of sentence F_0 contours in speech than either bottom- or topline-declination models. Huber [8] calculated a topline and a baseline.

In the current study, we will use a variant of the Lieberman method. We estimated the slope of declination by fitting an all-points regression line to the F_0 points (with Hertz values transformed into semitones), in the phrases and utterances determined by the mean scores of the nine transcribers. The method is visualized in figure 1.

Time-dependencies were investigated by calculating correlations between declination slopes (in semitones/second) and the length (in seconds) of phrases and utterances. The reason for doing so is because of claims in the literature (e.g. [6]) that the rate of declination is dependent on the length of a unit, the effect being that the slope of declination decreases as duration increases.

Unit	Slope	Duration
Phrases		
spon	-0.54 (1.11)	3.87 (2.59)
read	-1.67 (1.13)	2.70 (1.17)
Utterances		
spon	-0.44 (0.44)	8.06 (4.56)
read	-0.88 (0.60)	7.07 (2.47)

Table 1: Means (and respective standard deviations) of slope (semitones/second), and duration (seconds) of phrases and utterances in the spontaneous and read monologues.

Next to these global measures, we also determined local features of declination by investigating resets in topline declination. This was determined by considering the relationship (expressed in semitones) between two consecutive F_0 peaks located at either side of a phrase or an utterance boundary.

3. RESULTS

3.1. General

Before discussing the specific results related to aspects of declination, we will give some general prosodic findings. We found, first, that mean F_0 was higher in read aloud (136 Hz) than in the spontaneous (107 Hz) mode. This finding is in agreement with most earlier studies, but the fact that F_0 in spontaneous speech is lower is certainly not generalizable to all other types of spontaneous speech. Blaauw [1], e.g., claims that this effect may be due to the fact that most types of spontaneous speech are less careful speaking styles than most read-aloud speech, which can have its impact on average pitch.

As to the size of the units, it appears that the two speaking styles have phrases and utterances that are comparable in length (read: 19 words/utterance and 8 words/phrase; spontaneous: 16 words/utterance and 8 words/phrase).

3.2. Specific

Slope Using the regression line procedure sketched above, we found the results that are given in table 1. Generally, the analysis revealed negative slopes in phrases in both the read and spontaneous speech, but the slopes were significantly steeper in the former than in the latter ($F_{(1,59)}=15.557$, $p<0.001$), as was the difference in duration ($F_{(1,59)}=5.263$, $p<0.05$). A similar picture emerges for the utterances that appear to be significantly different in terms of slope ($F_{(1,26)}=5.049$, $p<0.05$), although duration appears not to be different ($F_{(1,26)}=0.460$, n.s.).

Time-dependency The data in table 1 above already suggest that a time-dependency can be observed in our speech materials, since utterances appear to be flatter than phrases, although this difference is stronger in the read data. To study

this more elaborately, we calculated correlation coefficients between the length of phrases and utterances and the corresponding slopes. The significance of the correlations was tested using Fisher’s r to z . Results are given in table 2. The analysis revealed significant correlations for utterances in both read and spontaneous speech (read: $p=0.0016$; spontaneous: $p=0.0229$). For phrases, only the correlation for read speech was significant (read: $p=0.0153$; spontaneous: $p=0.1958$). Furthermore, we found higher correlations in read than in spontaneous speech and higher correlations for utterances than for phrases.

Unit	r
Phrases	
spon	0.244
read	0.429
Utterances	
spon	0.559
read	0.782

Table 2: Correlation between declination and duration of phrases and utterances in the spontaneous and read monologues.

Reset The results for resets at phrase and utterance boundaries in the two speaking styles are given in table 3. The table reveals that there is indeed a tendency for F_0 to reset at the boundaries of linguistic units, but there appears to be a difference in degree. The analysis brought to light that read speech in general exhibits the strongest resetting as compared to spontaneous speech. Read utterances differ significantly from the spontaneous ones ($F_{(1,24)}=5.89$, $p<0.05$), whereas the difference between read and spontaneous phrases is not significant ($F_{(1,24)}=0.61$, n.s.). Also, analogous to the F_0 minimum values and the pause results in similar data ([13]), the difference in boundary strength between phrases and utterances is to some extent reflected in the amount of resetting. Utterances differ significantly from phrases in the read materials ($F_{(1,27)}=10.78$, $p<0.01$), whereas in the spontaneous monologues the difference in reset between those units is only a non-significant trend ($F_{(1,27)}=3.30$, $p=0.0803$).

4. DISCUSSION

Based on this study we conclude that there are, in effect, prosodic differences between the two speaking styles investigated, but that these differences should be described in quantitative rather than qualitative terms. That is, both read-aloud and spontaneous speech are characterized by negative slopes of declination and by resetting, and they both exhibit a time-dependency. Nevertheless, differences show up between the two speaking styles in the degree to which they exhibit these patterns.

A comparison of spontaneous and read speech in terms of

Unit	mean reset
Phrases	
spon	0.06 (2.84)
read	0.92 (3.22)
Utterances	
spon	2.06 (3.05)
read	5.36 (3.89)

Table 3: Mean reset (and respective standard deviations; in semitones) at phrase and utterance boundaries in the spontaneous and read monologues.

global prosodic features may have psycholinguistic relevance. In particular, it may shed light on such questions as preplanning and look-ahead. Most evidence to date seems to point to the fact that a speaker needs only a little amount of look-ahead to be able to adequately produce sentences ([10]). The dependency between the length of a unit and the amount of F_0 fall somewhat contradicts this finding, since this result suggests that a speaker may ‘overlook’ a whole utterance. At the same time, it appears that this result is dependent on the speaking style investigated. The finding that this dependency is clearer in read-aloud speech may be due to the fact that in this style speakers literally have an overview of the length of an utterance.

ACKNOWLEDGEMENTS

Part of this study was undertaken within the framework of the Language Technology Program supported jointly by the Swedish Council for Research in the Humanities and Social Sciences and the Swedish National Board for Industrial and Technical Development. Marc Swerts is also affiliated with the University of Antwerp (UIA). Part of the research was done while he was a visitor at the phonetics department of Umeå University.

REFERENCES

1. E. Blaauw (1995), On the perceptual classification of spontaneous and read speech. OTS Dissertations series. Research Institute for Language and Speech, Utrecht University.
2. G. Bruce (1982a), “Developing the Swedish intonation model,” Working papers 22, Lund University, pp. 51–116.
3. G. Bruce (1982b), “Textual aspects of prosody in Swedish,” *Phonetica* 39, pp. 274–287.
4. A. Cruttenden (1986), *Intonation*. Cambridge: Cambridge University Press.
5. J. ’t Hart (1986), “Declination has not been defeated – A reply to Lieberman *et al.*,” *Journal of the Acoustical society of America* 80, pp. 1838–1841.
6. J. ’t Hart, R. Collier and A. Cohen (1990), *A perceptual study of intonation*. Cambridge: Cambridge University Press.
7. M. Horne, E. Strangert and M. Heldner (1995), “Prosodic boundary strength in Swedish: Final lengthening and silent interval duration,” *Proc. XIIIth International Congress of Phonetic Sciences*, Stockholm, Sweden, Vol. 1, pp. 170–173.
8. D. Huber (1989), “A statistical approach to the segmentation and broad classification into phrase-sized information units,” *IEEE ICASSP-89*, Glasgow, Vol. 1, pp. 601–603.
9. R.D. Ladd (1984), “Declination: a review and some hypotheses,” *Phonology yearbook* 1, 53–74.
10. W. Levelt (1989), *Speaking: from intention to articulation*. Cambridge: Cambridge university press.
11. P. Lieberman *et al.* (1985), “Measures of the sentence intonation of read and spontaneous speech in American English,” *Journal of the Acoustical Society of America* 77, 649–657.
12. J.R. de Pijper and A.A. Sanderman (1994), “On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues,” *Journal of the Acoustical Society of America* 96, pp. 2037–2047.
13. E. Strangert (1993), “Speaking style and pausing,” *PHONUM* 2, pp. 121–137.
14. E. Strangert and M. Heldner (1995), “Labelling of boundaries and prominences by phonetically experienced and non-experienced transcribers,” *PHONUM* 3, pp. 85–109.
15. M. Swerts and R. Gelyukens (1994), “Prosody as a marker of information flow in spontaneous speech,” *Language and Speech* 37 (1), 21–43.
16. N. Umeda (1982), “ F_0 declination is situation dependent,” *Journal of Phonetics* 10, 279–290.