

# Prosodic signalling of (un)expected information in South Swedish – An interactive manipulation experiment

Gilbert Ambrazaitis

Linguistics and Phonetics, Centre for Languages and Literature, Lund University

Gilbert.Ambrazaitis@ling.lu.se

## Abstract

Starting from the German pitch peak timing categories and their communicative functions, it is asked how these functions would be expressed in South Swedish. The aim is to get a first impression as regards potentially relevant prosodic parameters associated with the expression of expected vs. unexpected information in South Swedish. For that, an interactive manipulation experiment is conducted, where subjects manipulate the pitch contour and duration of monosyllabic test utterances until the sound output adequately represents a given communicative function. Swedish has a tonal word accent distinction, and all test words have accent 1, normally produced with an early pitch fall. It is thus hypothesized that in South Swedish, expected vs. unexpected information will not be expressed through a different pitch peak timing, as in German. The results indeed clearly hint at unexpected information being signalled by means of a higher, rather than a later pitch peak.

## 1. Introduction

This study is part of a comparative investigation of intonational substance–function relations in German and Swedish. For German, Kohler has established a three-fold distinction of pitch accent timing: *early* vs. *medial* vs. *late peak* [1, 2]. The communicative functions associated with these peak categories can be summarized as ‘knowing’ (early), ‘realizing’ (medial), and ‘realizing in contrast to one’s expectation’ (late) [3]. In a recent paper, Kohler argues for a fourth category, a *late-medial* peak [3], which is used to signal ‘contrast to one’s expectation’ as a matter-of-fact statement, while the late peak adds a personal expressive evaluation of that contrast. The relationship between a late pitch timing and the expression of contrast to one’s expectation is not, however, restricted to German, cf. Pierrehumbert and Steele for American English [4].

The present study represents a first step in investigating how the functions associated with the German pitch peak timing categories are expressed in South Swedish. It is assumed that, in German, it is hardly possible to relate the phonetic parameter *timing* to only one functional parameter *degree of ‘knowing’*. Kohler [3] has shown that the four hypothesized peak timing categories are better distinguished with reference to more than just a single scale. However, it has not yet been clarified which scales or functional aspects, if any, may be regarded as more or less basic. A functional parameter labelled ‘expected/ unexpected’ is chosen to start with here, which rather intuitively appears to represent a potentially important basic functional aspect associated with the German pitch peaks.

Swedish has a word accent distinction, which – in South Swedish – is manifested through a different timing of the pitch accent fall: In single-word utterances, accent 1 is realized with

an early fall (the F0 peak being located around the stressed-syllable onset), and accent 2 with a late fall (F0 peak located around the stressed-syllable offset) [5]. The Swedish word accent distinction, however, exists only for words with non-final stress; words with final stress bear accent 1.

The aim of this study is to get a preliminary impression about how expected/ unexpected information is manifested prosodically in South Swedish single-word utterances consisting of monosyllabic words. Are expected and unexpected information – as in German – signalled through an earlier, or later timing, respectively, compared to a medial position associated with ‘neutral’ new information? On the basis of the summary above, two competing hypotheses can be formulated:

- H1** All monosyllabic words bear accent 1, and thus a monosyllabic utterance exhibits an early pitch fall. The functional parameter ‘expected/ unexpected’ will *not* affect the timing of the pitch accent in South Swedish monosyllabic utterances in the same manner as in German and English. Timing is blocked – at least to a certain degree – due to word accent contrast preservation, cf. [6]. The functional parameter ‘expected/ unexpected’ will instead affect other properties of the prosodic substance, e.g. peak height, which may work as a substitute for timing in some contexts, cf. [6].
- H2** The functional parameter ‘expected/ unexpected’ will affect the timing of the pitch accent in South Swedish monosyllabic utterances in a manner comparable to German. Since there is no word accent contrast for monosyllabic words, contrast preservation is irrelevant and a later timing of the pitch fall is not blocked as in H1. Empirical support for H2 is provided by a study on Central Swedish [7], which has shown that peak timing is a strong perceptual cue in connection with the interpretation of clarification ellipses.

The elicitation of the functional contrast by means of constructing test utterances or dialogues, which would be read by subjects, was deemed to be a rather difficult method. If adopted, subjects would at least have to be instructed to act out the dialogue, which would at best end at *acted* – rather than *elicited* – expressions of expected or unexpected information. Ideally, this functional contrast should be analyzed in spontaneous speech in natural communicative settings. For the present study instead, another rather simple experimental set-up was used, referred to as an interactive manipulation experiment.

## 2. Method

The general idea was to let subjects adjust a number of acoustic parameters themselves, until the so-created prosodic pattern satisfactorily represents a given function (e.g. ‘unexpected’). The

software Praat [8] was used both for recording the test material, preparing the material for the experiment, and as the manipulation tool for the subjects. In a future step, when potentially relevant acoustic parameters have been identified, one could consider building a user interface where specific acoustic parameters can be adjusted by the use of scrollbars. This would have the advantage that the subjects would not be confronted with the actual parameter curves. Such an experimental set-up has been used e.g. by Bruce and Granström [9]. However, since the purpose of this study was to get a first impression as regards the relevant parameters associated with ‘expected/ unexpected’, it was deemed important not to constrain the subject’s manipulation facilities to any considerable extent at this point.

### 2.1. Material

Three single-word utterances containing a monosyllabic accent 1 word each – *blå* (blue), *gul* (yellow), and *röd* (red) – were recorded by a 35 year-old male speaker of South Swedish with monotonous pitch. Since the original productions contained extremely long vowels (450–500ms), the material was further manipulated by shortening the vowels down to approximately 50% of the original duration.

### 2.2. Pilot test

The concept of more or less expected information is rather difficult to explain to naive subjects. Therefore, a pilot test with more easily distinguishable communicative functions (‘definite’, ‘surprised’ and ‘questioning’) was run to test the general applicability of the experimental set-up. A group of 22 students aged 17–18, working in groups of 2–3, took part as subjects.

Three versions of each test word were presented as objects in the main menu of Praat. These objects were labelled e.g. *blå-förvånad* (blue-surprised), indicating the test word and the target communicative function. The manipulation files were prepared such that the pitch contour was stylized, with a varying number of pitch points (3–4). After the experiment the subjects judged the quality of their manipulations on 5-grade scales, and gave comments on the experiment, partly through setting a cross on 5-grade scales (fun factor of the experiment; technical difficulty handling the program; general difficulty of the task), and partly in the form of open comments.

The main results can be summarized as follows. The subjects judged the experiment as being fun to do, the handling of Praat as medium hard to easy, and the general task as medium to difficult. Three open comments were collected, two of which indicated that it would be better to have the possibility to also alter the duration of the test word. The subjects tended to be satisfied with their manipulations (+1.25 on a scale ranging from -2 to +2, averaged over all nine files and ten subject groups).

The pitch manipulations produced by the subjects varied largely with respect to the precise setting of the pitch points, the pitch range, the number of pitch points used and the resulting sound quality. However, as regards general aspects of the pitch contour shape, or the direction of pitch movement, the results were rather robust. For the function ‘definite’ 17 manipulations in total were collected (from three test words). One would hypothesize the pitch contour not to be rising, and indeed, only 2 rising contours were found. For ‘questioning’ one would expect the pitch contour more likely to be rising, and 13 out of 15 manipulations in total indeed had a rising pattern. For ‘surprised’, the results were more complicated. It is concluded from the pilot test that there are no general practical problems with the applicability of the experimental set-up.

### 2.3. Subjects

Six native speakers of South Swedish, aged 30–58, four male, two female, volunteered as subjects. They are referred to by S1–S6, S1 and S2 being the females. All are researchers or PhD students at the Centre for Languages and Literature, Lund University, but all were naive with respect to the purpose of the experiment. S4 was the speaker of the test material.

### 2.4. Procedure

The subjects received instruction in written form. They were given three sheets: an introduction sheet, an instructions sheet, and a working sheet. On the introduction sheet, the following situational setting was described: *Two friends are having a small chat. A: "By the way, Lasse has finally bought a new car!" B: "Really! It's high time! So what colour did he choose?" A: "Blue" (or "Yellow" or "Red").* It was then explained that A can utter the test word (e.g. *blue*) with a variety of melodies, and that these melodies can be interpreted in different ways. Three such possible interpretations were then explained by means of paraphrasing what A means when he says “*Blue*”: ‘*väntat*’ (‘expected’), ‘*oväntat*’ (‘unexpected’), and ‘*neutralt*’ (‘neutral’). The subject’s task was then described, and can be summarized as follows: Manipulate the melody and the duration of the test word yourself, until it sounds appropriate for a given context (e.g. ‘expected’). On the instructions sheet, the subjects were informed about how to handle the relevant Praat functions.

The test material was presented in the form of nine manipulation objects in Praat’s object list, labelled e.g. *röd-väntat* (*red-expected*). All nine files contained an equal number of pitch points (5), and the subjects were *not* allowed to remove or add any pitch points. The five pitch points were set in approximately equal temporal intervals throughout the sonorant part of the word (e.g. [lo:] in *blå*), stylizing the original monotonous contour at a medium pitch level. The F0 range was restricted by resetting the Praat parameters for upper and lower limits to 50Hz and 250Hz, respectively. The subjects were allowed to adjust all five pitch points along both axes. Additionally, they were allowed to manipulate the duration of the rime. For that, one duration point was set near the end of the onset consonant (or cluster), and a second point closely after the vowel onset. The subjects were allowed to adjust the second point only.

The subjects were instructed to keep in mind the following principles while working on the manipulations: Try to avoid creating a computer voice; imagine how you would have said the word and try to replicate the result in your manipulation. As in the pilot test, the subjects were asked to judge their manipulations on the working sheet, and to comment on the experiment. The subjects did the experiment one by one. They were instructed to keep within 30 minutes, if possible, but they were allowed to take the time they needed. The actual duration of a session varied between 30 and 60 minutes, mostly depending on the subject’s prior experience with Praat.

## 3. Results

The subjects tended to be satisfied with their manipulations (mean judgments per subject vary between 0 and 1.6 on a scale from -2 to +2). The technical handling of Praat was judged as easy (rating 1 or 2) by four subjects, neutral by one and difficult (-1) by another subject. The general difficulty of the task was rated on two separate scales (ranging from -2 to +2): (a) difficulty to imagine a suitable melody, and (b) difficulty to in-

teractively adjust the parameter points and to judge the quality of the synthesis. Both (a) and (b) were rated as slightly ‘difficult’ (a: -0.5, b: -0.2, averaged over all six subjects), but the actual ratings ranged from -2 to +1 for (a) and from -2 to +2 for (b). Four subjects gave an open comment on the experimental set-up, and the most important points can be summarized as follows:

1. It was difficult to distinguish between ‘neutral’ and ‘expected’ (S1), but:
2. It was difficult to distinguish between ‘expected’ and ‘unexpected’, because ‘expected’ can sound ironic (S3).
3. More pitch points should have been available (S3), but:
4. Three or four pitch points would have been enough (S4).

Considering the aim of this study, the experimental set-up, the number of subjects, and the nature of the raised data, the results are presented first of all in terms of qualitative judgments. The general tendency seen in the data is that duration was hardly used to distinguish between the communicative functions, but pitch manipulation was used by most of the subjects to differentiate between all three functions. However, only ‘unexpected’ appeared to have a rather clear prosodic expression, namely a higher pitch peak, throughout the subjects. The other two functions showed different manifestations, if any, from subject to subject. Therefore, the remaining sections will concentrate on the difference between ‘unexpected’ on the one hand, and ‘neutral/ expected’ on the other hand.

### 3.1. Duration manipulation

S6 did not make any duration manipulations, and S3 only decreased the duration for one file (*blue-neutral*). The other four subjects used the duration manipulation quite extensively. S4 (= the speaker) systematically decreased the duration of the rime in every file by approximately 35% on the average. S1 systematically *increased* the duration for ‘unexpected’, but *decreased* the duration in most of the other cases. S2 and S5 decreased the duration in some cases, increased it in other cases, or left the duration unchanged, without showing any systematic pattern. In summary, one out of six subjects appears to have made a systematic duration distinction with respect to the target functions (cf. the lengthening for ‘unexpected’ by S1).

### 3.2. Pitch manipulation

As in the pilot test, the resulting pitch contours varied greatly in detail. However, there are some general tendencies. With some exceptions, the subjects created only falling intonation patterns. One exception is S6, who created some form of falling-rising pattern for 8 of the 9 combinations of test word and function, and therefore will not be considered further. Another exception is S2, who created falling patterns for ‘neutral’ and ‘expected’, but a rising pattern for ‘unexpected’ on all three test words. A third exception is S5, who created a more or less monotonous pattern for ‘neutral’.

The subjects tended to leave the first pitch point unchanged, and to adjust the remaining pitch points to a greater extent along the vertical than along the horizontal axis.

#### 3.2.1. Pitch timing

For the falling contours created by S1–S5, the timing of the pitch peak was estimated by measuring the temporal distance between the vowel onset and the F0 maximum. Since the duration manipulations affect the pitch peak timing, the measurements were undertaken in a resynthesized sound file, rather than the original manipulation file. In many cases, the subjects had

created a high plateau; in other cases, a slight rise preceded the F0 maximum, which was visible in the contour, but perceptually equivalent to a high plateau. In such cases lacking a clear F0 peak, the temporal distance between the vowel onset and the onset of the high-pitched portion was measured. Since only general trends are of interest here, all timing values were rounded off to the nearest 5ms. Table 1 displays the results.

Table 1: *Peak timing measured from vowel onset in [ms] for S1–S5. No values for rising (r) and monotonous (m) patterns.*

File	S1	S2	S3	S4	S5
blue-neutral	0	0	0	0	(m)
blue-expected	15	0	0	0	0
blue-unexpected	15	(r)	20	0	20
yellow-neutral	0	90	95	30	(m)
yellow-expected	90	170	95	0	65
yellow-unexpected	90	(r)	100	60	115
red-neutral	50	65	50	20	(m)
red-expected	50	195	65	0	55
red-unexpected	75	(r)	50	0	65

The timing of the F0 peak with respect to the vowel onset varies between 0 and 195ms, i.e. a quite late timing was created in some cases. However, there is no regular pattern observable in the timing differences with respect to the communicative functions. Moreover, for the test word *blâ* (*blue*) the F0 maximum is found near the vowel onset for all functions and all subjects. It has to be pointed out that *blâ* is the only test word with a sonorant consonant within the syllable onset (/r/ in *röd* is realized as a uvular fricative). It is thus likely that the timing variation found in the data is largely due to an artifact of the test material. Recall that the pitch points were set throughout the sonorant portion of the test word. That is, only *blâ* was provided with a pitch point clearly preceding the vowel onset. Consequently, the general tendencies in the subject’s adjustment strategies (see 3.2 above) have resulted in later timing of the pitch peak in *gul* and *röd* as compared to *blâ*.

#### 3.2.2. Pitch height

The maximum F0 values for the purely falling contours created by S1–S5 were measured (rounded off to the nearest 5Hz), and the results are displayed in table 2. With few exceptions, all subjects have created a higher peak for ‘unexpected’ as compared to the other two functions. This is most pronounced for S4, the speaker of the test material. Figure 1 contains the contours created by S4 for the test word *blâ* as an example.

Table 2: *Peak height in [Hz] for S1–S5. No values for rising (r) and monotonous (m) patterns.*

File	S1	S2	S3	S4	S5
blue-neutral	125	170	150	130	(m)
blue-expected	130	200	145	115	150
blue-unexpected	135	(r)	230	225	210
yellow-neutral	125	145	180	130	(m)
yellow-expected	135	175	155	115	145
yellow-unexpected	135	(r)	190	225	225
red-neutral	125	160	195	130	(m)
red-expected	130	155	160	115	125
red-unexpected	150	(r)	210	225	200

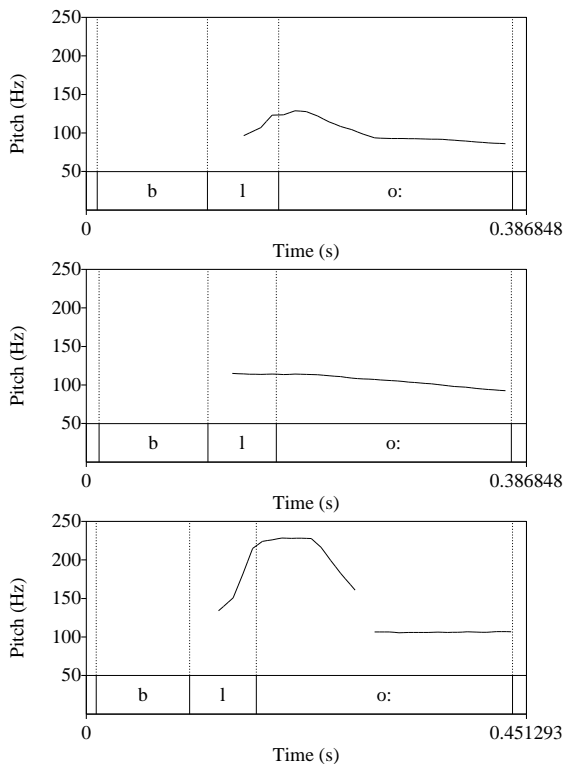


Figure 1: Results of the manipulations for test word blue by S4. Top to bottom: neutral, expected, unexpected.

#### 4. Discussion

The design of the present experiment needs refinement in some aspects. First of all, the communicative functions under investigation have to be defined for the subjects more carefully, and perhaps different situational contexts should be used for that purpose. For example, the label ‘neutral’ should probably be avoided, since subjects had different kinds of problems with ‘neutral’, cf. the open comment by S1, or the monotonous pitch contours by S5, respectively.

Nevertheless, the present study has provided a preliminary insight as regards potentially relevant prosodic parameters associated with the expression of ‘expected/ unexpected’ information in South Swedish. The duration of the rime was systematically manipulated by only one subject with respect to the communicative functions. Durational factors are thus perhaps not too relevant. Furthermore, the data do by no means support pitch peak timing as a relevant parameter, since no systematic variation of timing with respect to the communicative functions was created by any of the subjects. This would provide support for H1, cf. Introduction. However, this support is somewhat constrained by the fact that a clearly early accent 1 fall could not be attested in every instance. The precise timing created by the subjects is probably due to a methodological artifact. That is, a clear case for contrast preservation cannot be attested by the present data either.

Pitch peak height, however, was manipulated very systematically for falling contours by five subjects, since ‘unexpected’ received a higher peak than ‘neutral’ and ‘expected’. This result also supports H1. Therefore, future research will examine

the association of pitch height and expression of unexpected vs. expected information more systematically. Three types of data are being considered for that purpose. First, real production data should be examined. For that, an attempt has to be made to elicit production data, although deemed to be difficult, and spontaneous speech data should also be included. Second, the effect of pitch height on the perception of (un)expected information will be examined. Third, further interactive manipulation experiments should be run, with a considerable restriction of the subject’s manipulation possibilities, comparable to [9]. That is, the subjects could be confronted with a scroll bar, instead of a parameter curve, for adjusting the parameter pitch height.

The experiment provided no clear insight as regards the expression of ‘expected’ information as compared to a ‘neutral’ answer. However, the results gave some hints as regards potentially relevant acoustic parameters. For example, the contours for ‘neutral’ and ‘expected’ in figure 1 differ largely with respect to the shape of the F0 fall, which is much more abrupt for ‘neutral’. This internal timing of pitch peaks has also proved important as regards the German pitch peak categories, cf. [10].

Future research should therefore include other parameters than just pitch peak height; timing variables might still have some relevance. Furthermore, the investigation has to be extended to longer words, including words with accent 2, and longer utterances. Finally, it would be interesting to run a similar experiment with German material – where timing can be expected to be a relevant parameter – to test whether German subjects would manipulate the timing of the pitch peak. If they do not, the applicability of the method has to be reconsidered.

#### 5. Acknowledgments

My thanks go to the 22 plus 6 subjects, and especially S4 for recording the material, and Gösta Bruce, Merle Horne, and Oliver Niebuhr for valuable comments.

#### 6. References

- [1] Kohler, K., 1987. Categorical pitch perception. *11th ICPHS*, Tallinn, 331-333.
- [2] Kohler, K., 1991. Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics. *AIPUK 25*, Kiel, 115-185.
- [3] Kohler, K., 2005. Timing and communicative functions of pitch contours. *Phonetica*, 62(2-4), 88-105.
- [4] Pierrehumbert, J.B.; Steele, S.A., 1989. Categories of tonal alignment in English. *Phonetica*, 46, 181-196.
- [5] Bruce, G.; Gårding, E., 1978. A prosodic typology for Swedish dialects. *NORDIC PROSODY*, Lund, 219-228.
- [6] Gussenhoven, C., 2004. *The phonology of tone and intonation*. Cambridge: CUP.
- [7] Edlund, J.; House, D.; Skantze, G., 2005. The effects of prosodic features on the interpretation of clarification ellipses. *INTERSPEECH*, Lisbon, 2389-2392.
- [8] Boersma, P.; Weenink, D. Praat: Doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>.
- [9] Bruce, G.; Granström, B., 1993. Prosodic modelling in Swedish speech synthesis. *Speech Communication*, 13, 63-73.
- [10] Niebuhr, O., 2003. Perceptual study of timing variables in f0 peaks. *15th ICPHS*, Barcelona, 1225-1228.