

# Emphasis by Pausing

Eva Strangert

Phonetics at Department of Philosophy and Linguistics

Umeå University, SE-901 87 Umeå, Sweden

E-mail: eva.strangert@ling.umu.se

## ABSTRACT

The study concerns how prosody, including pauses, is used to give extra weight to focused words when repeated with successively more emphasis. The analysis is based on the same three-word phrase produced in two different communicative contexts, one a dialogue interaction and one a reading-aloud situation. Results show greater adjustments in the dialogue context. When emphasis was increased, there were more pauses and longer word durations than when words were just read aloud. F0 maxima in the words, however, were less affected by emphasis. These findings are seen in the light of different demands of the two communicative situations as well as different production constraints on F0 and duration.

## 1 INTRODUCTION

The research reported on here is undertaken within the framework of the Swedish GROG project “Boundaries and groupings – the structuring of speech in different communicative situations” [5]. The study concerns how prosodic focus is signalled in Swedish when words are successively more emphasized.

Focal accent is the highest level of prominence in the Swedish intonation model [2]. Focal accent, or prosodic focus, is signalled primarily by a rise in F0, although recent studies point to strong effects of focus on duration, and to a lesser extent also on intensity and spectral emphasis [8]. In addition, pausing appears to be a means to focus words. Strangert [10] reported on pauses in professional news reading. The pauses occurred *after* grammatical function words and *before* semantically heavy words, thereby helping to give them extra emphasis.

A focused word may be more or less emphasized, that is, within the phonologic category of focus a continuous variation of emphasis can be assumed [3]. It is in this context, when a word is emphasized, that pauses seem to occur. In addition to using pauses for emphasis, speakers employ temporal adjustments. Carlson et al. [4] report emphasis data from a single speaker. Analysing several readings of the same utterance they found a regular increase of word duration and F0 peak values as a function of emphasis. Ericson and Lehiste [6] report on longer word durations in emphasized than in non-emphasized words. (In addition, the non-emphasized words became shorter in utterances with an emphasized word. That is,

there was a temporal rearrangement of all the words in the utterance.) Also, in an analysis of prosodic properties of contrastive as compared to presentational focus, Selkirk [9] observed a frequent use of pauses when speakers aimed at contrastiveness. Thus, speakers tend to use pauses and other temporal adjustments (as well as F0 adjustments) under special focus conditions. It may happen whenever speech is produced to be maximally clear and informative, and in particular in order to signal contrast. It is therefore particularly in natural dialogue that pausing as a means for emphasis and contrast can be expected.

The interaction dimension and the listener perspective is central here. Borrowing from the H&H theory [7], the communication situation depends on the speaker *and* the listener. Thus, when a communication goes wrong, it is up to the speaker to adapt to the situation. The means used to do so include a variety of features and not the least repetitions [11]. Bell and Gustafson [1] analysed repetitive utterances in computer-directed speech. When the system did not cope with the utterances directed to it, durations increased and words were hyperarticulated or contrastively focused. Furthermore, second and third repetitions very often contained inserted pauses between words.

The current work aims at describing the means used when focused words are more and more emphasized. In addition to temporal adjustments, including pause insertions, the analysis will also take F0 adjustments into account. Two parallel experimental situations are studied. In the first one, a semi-spontaneous interaction between a speaker and a listener, the speaker is forced by the listener to increase emphasis in three steps. In the second one, the speakers produce the same material trying to increase emphasis successively.

In doing this, the assumption is first, that the demands on the speakers will be different in the two situations; the speakers will have to make greater adjustments in the interaction situation. Secondly, on the basis of previous work, pause insertions as well as adjustments in duration and F0 can be expected. However, as a Swedish focused word already has a high F0 peak value, and given the range over which a person’s F0 can vary, additional increase due to emphasis should be restricted. Duration, and in particular pause insertions should be less constrained. Therefore, the means for achieving higher emphasis are expected to be primarily temporal.

## 2 METHOD

12 native speakers of Swedish participated in the study, 6 in the interaction and 6 in the reading experiment. All were students of about the same age, and all except one were women.

All 12 speakers produced the same NP-V-NP phrase, *Ellen säljer fällen* /sɛlˈjɛr ˈfɛllɛn/ 'Ellen is selling the rug'. The phrase was presented to the speakers before the experiment and they could read it on a sheet of paper placed in front of them during the whole recording session. The phrase was uttered with focus on each of the three words. Focus on each of the words in turn was elicited by the following questions (in English translation) posed by the experimenter:

*Ellen*: 'Who is selling the rug?'

*säljer*: 'What is Ellen doing?'

*fällen*: 'What is Ellen selling?'

For each of the three phrases, four tokens were produced with successively increasing emphasis on the focused word. 6 of the subjects uttered the phrases in a semi-spontaneous dialogue with the experimenter (the author) pretending to misunderstand the focused word and repeatedly asking for clarification [cf. 1995]. Thus, to induce increased emphasis the experimenter said (in English translation): 'I didn't get it. Could you repeat that again.' And then to induce even more emphasis, the experimenter used variants of: 'I am sorry, but still I can't get it. Could you repeat that again.' The elicitation procedure started from a low (reference) level and emphasis was increased in three steps on each of the focused words. The four levels will be referred to as e0-e3 (for *Ellen*), s0-s3 (for *säljer*) and f0-f3 (for *fällen*). In the reading experiment, the other 6 speakers read the same phrases (with focus position marked by underlining in the text) trying to increase emphasis successively in three steps from the first (reference) level.

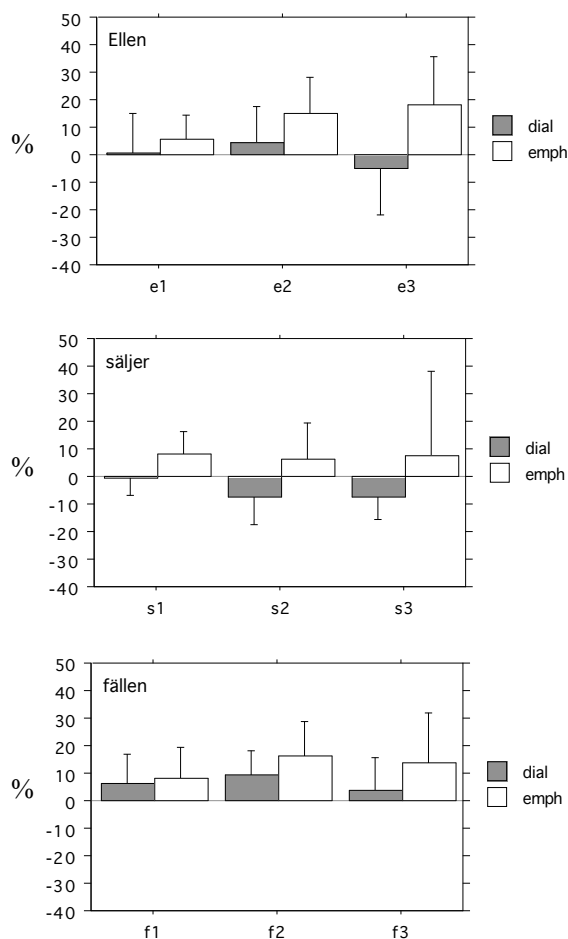
The recordings were made in a sound-treated recording studio using high quality professional equipment and were digitised at 16 kHz. Analysis was based on measurements made after segmentation and labelling using ESPSwaves+. Measurements for each phrase included F0 peak values (in Hz) and duration of the focused word (in msec). That is the analysis was restricted to the words in focus. The F0 and duration data were subsequently used for calculating the successive change from lower to higher levels (0-3) of emphasis for each speaker. The changes were expressed in percent with the lowest level (0) as the reference for both F0 maxima and word duration. Finally, in cases of a pause occurring before or after the word in focus, pause (= silent interval) duration (in msec) was measured.

## 3 RESULTS

### 3.1 F0 ADJUSTMENTS

Figure 1 (a-c) shows the change in F0 peak values for each

of the three words separately across the four levels of emphasis (e0-e3, s0-s3 and f0-f3. Data (means for the six speakers) are given as percentage changes at step 1, 2 and 3 relative to the references (e0, s0 and f0). Interspeaker variation is expressed by 95% confidence intervals.

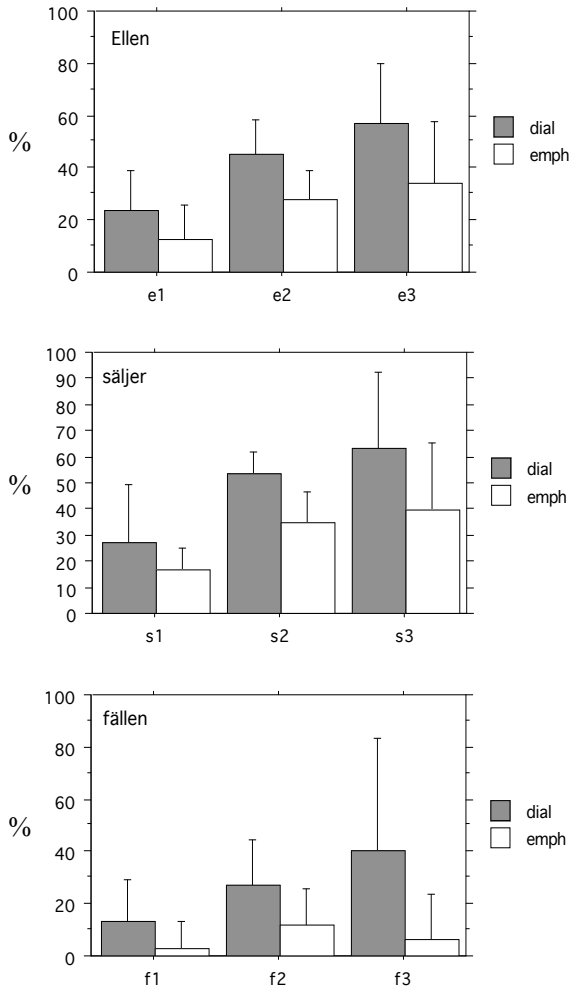


**Figure 1:** Mean changes of F0 peak value in focused words upon emphasis in three steps (e1, e2, e3 for *Ellen*; s1, s2, s3 for *säljer*; f1, f2, f3 for *fällen*) in percent relative to a low emphasis reference value (e0; s0; f0) for each word. Means for six speakers.

The adjustments in F0 appear as positive or negative changes relative to the reference level for each word. In the reading experiment, increased emphasis in general leads to increased mean F0 peaks. However, as indicated by the spread in the data, not all of the speakers increase F0 maximum step by step. A closer analysis also shows that in some cases there is even a lowering of the peak value. Furthermore, there is a difference between the words, *säljer* being the least affected by emphasis. Generally the mean increase in peak value centres around 5-15%. The adjustments are greater than in the dialogue experiment, in which the mean changes never exceed 10%. However, the changes in the dialogue are positive in some cases and negative in others. This is true for *Ellen* as well as *säljer*. Only in the last word, *fällen*, the change is generally positive.

### 3.2 ADJUSTMENTS OF WORD DURATION

Figure 2 (a-c) shows the change in word duration for each of the three words separately across the four levels of emphasis (e0-e3, s0-s3 and f0-f3). Data (means for the six speakers) are given as percentage changes relative to the references (e0, s0 and f0). Interspeaker variation is expressed by 95% confidence intervals.



**Figure 2:** Mean changes of word duration in focused words upon emphasis in three steps (e1, e2, e3 for *Ellen*; s1, s2, s3 for *säljer*; f1, f2, f3 for *fällen*) in % relative to a low emphasis reference value for each word (e0; s0; f0). Means for six speakers.

Word duration increases successively from low to high levels of emphasis in the reading as well as in the dialogue experiment. However the adjustments are greater in the dialogue experiment. While maximal increase is about 60-70% in the dialogue situation, the corresponding figure for the reading is about 40%. The means indicate an increase step by step as emphasis is increased. Though this is the general trend, the spread in the data indicate, and a more detailed analysis shows, that increases of duration do not occur at each step for each speaker. The trend of a stepwise increase is however similar for all the three words, although the last word, *fällen*, is the least affected with smaller changes in word duration.

### 3.3 PAUSE INSERTIONS

Data for the dialogue experiment are presented first. Table 1 shows the occurrence of pauses, their distribution relative to the focused word and pause durations. As can be seen, pauses occur frequently after focused *Ellen* and also before *säljer* but less frequently after the verb *säljer* and before the noun *fällen*. That is, most of the pauses (20 of the totally 30 pauses) occur between the first noun and the verb, at the NP-VP boundary. Pauses are very few at the lower levels of emphasis; they appear at the higher (2, 3) levels. Taking all 30 pauses into account, the durations range from 12 to 378 msec with a mean of 178 msec. Pause duration, furthermore, seems not to be affected by the level of emphasis. There is no trend of increased pause duration as a function of emphasis.

	<i>Ellen</i>	post	pre	<i>säljer</i>	post	pre	<i>fällen</i>
Sp1:0							
1							
2							
3			95				
Sp2:0		12					
1						84	
2		468	135				
3		500	146		257	58	
Sp3:0							
1							
2		136					
3		282	165				
Sp4:0			94				
1		205					
2		224	139		220		
3		201			130		
Sp5:0							
1		349	111		378		
2		90			184		
3		156	130				
Sp6:0							
1						61	
2		54					
3							165

**Table 1:** Distribution and durations of pauses (msec) before (pre) and after (post) focused *Ellen*, *säljer* and *fällen*. Individual data for 6 speakers.

In the reading experiment the total number of pauses is 13. However 6 of them were produced by the same speaker while the other pauses were spread between the different speakers. As in the dialogue, the majority (12) of the pauses occurred between *Ellen* and *säljer*, that is at the NP-VP boundary. The mean pause duration is 160 msec within a range from 18 to 581 msec. Of the pauses occurring at the NP-VP border, 4 were produced (after *Ellen*) by the same speaker. The durations of these pauses produced at each successive level of emphasis were 57 msec (e0), 238 msec (e1), 581 msec (e2) and 470 msec (e3). This speaker thus has a more or less successive increase from one step to another (except for the last step).

## 4 DISCUSSION AND CONCLUSIONS

It is obvious that the two different styles of speaking are characteristically different. Successively higher emphasis leads to adjustments in both situations, though to a lesser extent in reading aloud.

These differences most reasonably can be seen as resulting from the different demands on the speakers in the two situations. In reading aloud – without interaction with a (physically present) listener and thus without any feedback about communicative success or failure – the speaker has to adapt to what he/she *believes* is required in the specific situation. A dialogue situation, such as the one in the current study, is considerably more demanding. When forced by the listener to be more and more distinct, the speaker has to put in some extra effort. The means to do this are primarily temporal. The speakers lengthen the focused word much more than in the reading-aloud situation, and insert pauses much more frequently. From the lowest to the highest level of emphasis the mean word lengthening amounts to 40-60%, with the greatest lengthening in *Ellen* and *fällen*. Most of the pauses occur at the higher levels of emphasis. However, pause duration does generally, in contrast to word duration, not increase as a function of emphasis. The predominant pause position is *after* focused *Ellen* and *before* focused *säljer*. That is, the pauses occur at the boundary between the NP and V and less so between the V and second NP, see Table 1. This preponderance of pauses at the NP-VP boundary might indicate a possible link to syntax. Whatever the reason for the skewed distribution, however, pausing – in combination with increased word duration – appears as a powerful means used to make words extra prominent.

F0 adjustments in contrast, seem to be very restricted. (cf. [4], however, in which F0 increased substantially as a function of emphasis in reading). These differences between F0 and temporal cues most reasonably can be explained by different production constraints on duration (including pausing) and F0. Fewer constraints can be assumed for adjustments of duration and pause insertions than for F0 adjustments. A focused word already has a high F0 peak and given the range over which F0 can vary, an additional increase due to emphasis has to be restricted. However, the different patterns in the dialogue experiment (with a decrease in F0 peak values in two of the words) and the reading experiment (with positive changes) cannot be explained on the basis of the data presented. It may be that the speakers in the dialogue start out from a higher level and therefore cannot change F0 peaks any further. However, as the speakers in the dialogue were not the same as in the reading, this assumption cannot be checked against data. Finally, the adjustments in the dialogue point towards a possible trading relationship between F0 and duration, as F0 peak values tend to decrease, when word durations and pause insertions increase.

In conclusion, the different demands in the two communicative situations is manifested in different patterns of adjustments. The extra distinctiveness needed

in the dialogue interaction, which possibly cannot be achieved by F0 adjustments, is met by extreme temporal adjustments and a frequent use of pausing. In particular, pausing occurs at the higher levels of emphasis, when extraordinary high prominence is required.

## 5 ACKNOWLEDGEMENTS

Thierry Deschamps is gratefully acknowledged for assistance in the preparation of the acoustic data for analysis. The study was supported by a grant from the Swedish Research Council.

## REFERENCES

- [1] L. Bell and J. Gustafson, "Repetition and its phonetic realizations: investigating a Swedish database of spontaneous computer-directed speech," *Proc. ICPhS 99, San Francisco*, pp. 1221-1224, 1999.
- [2] G. Bruce, *Swedish Word Accents in Sentence Perspective*, Lund: CWK Gleerup.
- [3] G. Bruce, "Allmän och svensk prosodi," *Praktisk Lingvistik*, 16. Lund: Dept. of Linguistics, 1998.
- [4] R. Carlson, Y. Erikson, B. Granström, B. Lindblom and K. Rapp, "Neutral and emphatic stress patterns in Swedish," in *Speech Communication*, vol 2. Stockholm: Almqvist & Wiksell, 1975.
- [5] R. Carlson, B. Granström, M. Heldner, D. House, B. Megyesi, E. Strangert and M. Swerts, "Boundaries and groupings – the structuring of speech in different communicative situations: a description of the GROG project," *TMH-QPSR*, 44, pp. 65-90, 2002.
- [6] D. Ericson and I. Lehiste, "Contrastive emphasis in elicited dialogue: durational compensation," *Proc. ICPhS 95, Stockholm*, vol 4, pp. 353-355, 1995.
- [7] B. Lindblom, "Explaining phonetic variation: a sketch of the H&H theory," in *Speech Production and Speech Modelling*, W. J. Hardcastle and A. Marchal, Ed., pp 403-439. Dordrecht: Kluwer Academic Publ. 1990.
- [8] M. Heldner, "Focal accent – f0 movements and beyond," *PHONUM*, 8. Reports in Phonetics Umeå University, 2001.
- [9] E. Selkirk, Contrastive *FOCUS* vs. presentational *focus*: Prosodic evidence from right node raising in English," *Proc. Speech Prosody2002*, pp. 643-6, 2002.
- [10] E. Strangert, "Phonetic characteristics of professional news reading," *PERILUS*, XIII, pp 39-42. Institute of Linguistics, Stockholm, 1991.
- [11] M. Swerts, H. Koiso, A. Shimojima and Y. Katagiri, "On the different functions of repetitive utterances," *Proc. ICSLP '98, Sydney*, Vol 2, pp. 483-486. 1998.