BMBF

Verbmobil
Verbundvorhaben

Distinction between 'normal' and 'contrastive/emphatic' Focus

Anja Elsner

IKP Universität Bonn

Vm Report 167
August 1996

August 1996

Anja Elsner

Institut für Kommunikationsforschung und Phonetik
Universität Bonn
Poppelsdorfer Allee 47
53115 Bonn

Tel.: (0228) 7356 - 53
Fax: (0228) 7356 - 39
e-mail: {ape}@ikp.uni-bonn.de

Gehört zum Antragsabschnitt: 15.5

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 101 D 08 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei der Autorin.

to appear: ICSLP 1996, Philadelphia

DISTINCTION BETWEEN 'NORMAL' AND 'CONTRASTIVE/EMPHATIC' FOCUS

Anja (Petzold) Elsner
e-mail: ape@ikp.uni-bonn.de

Institut für Kommunikationsforschung und Phonetik (IKP), University of Bonn,
Poppelsdorfer Allee 47, 53115 Bonn, Germany

ABSTRACT

A method to extract and classify focus accents has been developed. It works for German spontaneous speech. The method tries to distinguish 'normal' and 'contrastive/emphatic' focus accents using phrase boundaries. It was found that contrastive/emphatic accents tend to have greater distances to phrase boundaries than normal focus accents. Moreover, for contrastive/emphatic accents there was found a much steeper F_0 rise for accents with a rather high distance from the next phrase boundary.

1. INTRODUCTION

In the German research project VERBMOBIL several modules work together in recognizing speech from spontaneous dialogues. Further processing includes translation and synthesis of these dialogues in this application. It is important for translation and synthesis to have additional prosodic information which otherwise would have to be derived from the linguistic context. In many cases it is even impossible to recognize special emphasis intended by speakers in spoken dialogues without prosody.

Let us consider the following examples (focussed parts emphasized):

(1a) In **der** *Woche kann ich nicht.*

(In this week it's impossible for me. [but perhaps in another week])

(1b) In der **Woche** *kann ich nicht.*

(During the week it's impossible for me. [but perhaps on the weekend])

In this case linguistic analysis has no chance to find the correct meaning without prosodic informatio. Nevertheless it is likely that a speech processing system takes the second version (1b) as standard interpretation because 'Woche' (week) is a content word. The other reason is that we normally would expect a phrase boundary after 'Woche' - we will come back to this later (see Section 3).

(2a) Ich könnte um elf noch einen **Termin** *reinschieben*

(I could insert another **date** at eleven o'clock.)

(2b) Ich könnte um **elf** *noch einen Termin reinschieben*

(I could insert another date at **eleven** o'clock.)

At first sight we have no meaning difference in this example. However, in (2b) we could suppose that the *only* possible time for the speaker would be eleven o'clock, while this question is left open in (2a). Thus, for translation or synthesis of this sentence we need prosodic information to reproduce the original intention of the speaker even if it implies only a small shift in meaning.

2. FOCUS RECOGNITION

Starting point for this investigation is an already existing algorithm for focus recognition. Focus is defined here as the semantically most important part of an utterance, which is in general marked by prosodic means. The focus accents reflect the intention of the speaker to mark those parts of a sentence which he feels to be important. Normally these are content words. Nevertheless, in special contrastive/emphatic aspects it is also possible to put a focus accent on a function word (see example (1a)).

Our method for focus recognition is as follows [1]: The algorithm tries to solve focus recognition by global description of the utterance contour, in a first approach represented by the fundamental frequency F_0. Investigations of Swedish spontaneous speech [2] have shown that declination can be controlled by the focal accent: It was found that in pre-focal position there is no downstepping, but after a focal accent downstepping is significant and characteristic.

To examine this feature in German spontaneous speech, the reference line was computed as follows: First the F_0 contour was postprocessed by a special smoothing algorithm described in [3]. (Without smoothing results get worse by about 7 %.)

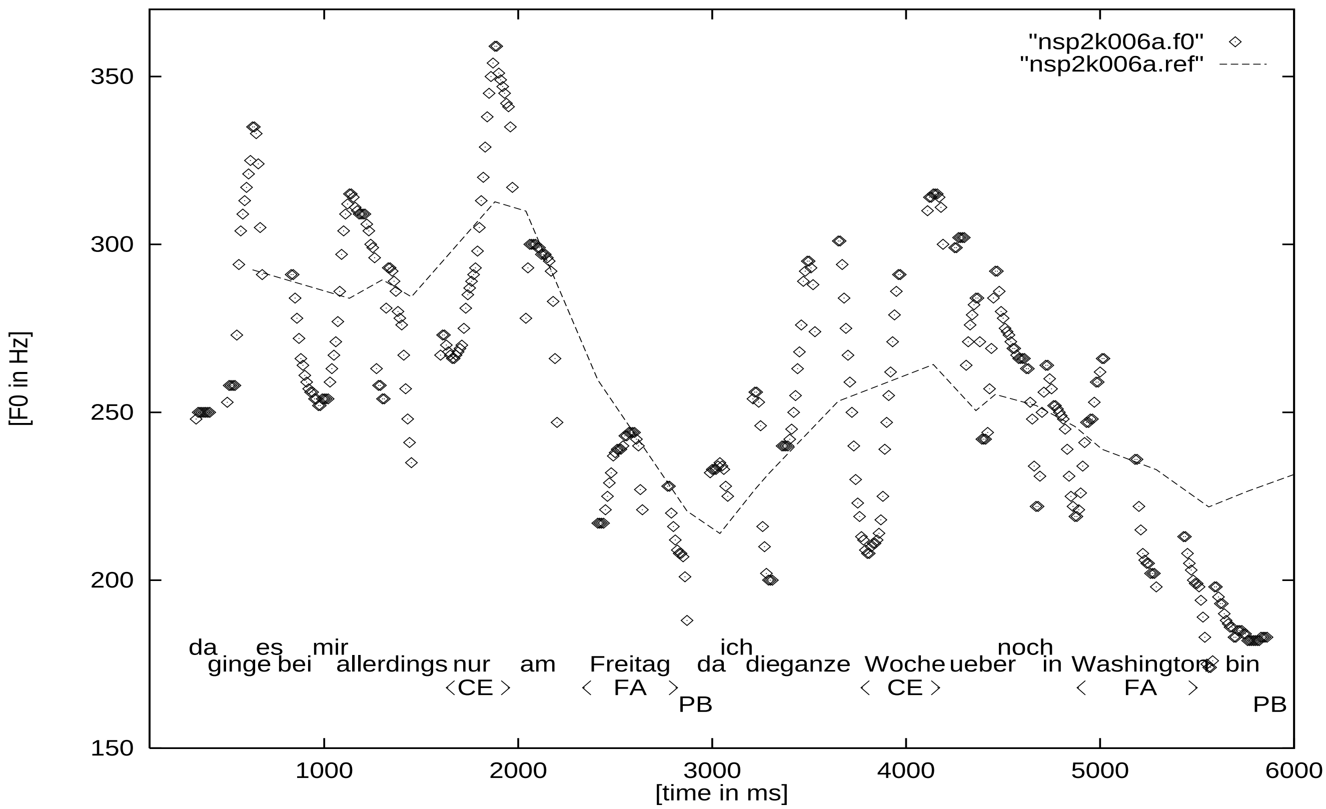

Figure 1: Utterance extracted from a dialogue with reference line and labelled focus (FA), contrastive/emphatic accent (CE) and phrase boundary (PB).
("For me it's only possible on friday because I'll be in Washington all over the week")

In a second step significant maxima and minima in a window of 90 ms size were detected. The average values between the maximum and minimum lines yield the global reference line.

According to [2] the focus must be in the area of the steepest fall in the F_0 course. Therefore the points with the highest negative gradient were determined first in each utterance. To determine the position of the focus the nearest maximum in this region has been used as approximation.

The global recognition rate is 78.5 % and the mean recognition rate is 66.6 %. The recognition rate for focus areas (45.8 %) is significantly worse than for nonfocus areas (87.5 %), i. e. there are far more deletions than insertions. Only a minority of the frames fall within focused regions (18.5 %). In a collaboration with other modules it is better that a focus remains undetected - false alarms may cause more problems.

In **Figure 1** we see an example of the focus detection algorithm. Following the reference line (streaked line) the algorithm detects 'nur' (only) and 'Woche' (week) as focus accents; the 'default' focus accents (FA) close to the phrase boundary (PB) remain undetected.

While recognition rates are acceptable, the computation of the reference line is sometimes incorrect or the focus is located in a question or continuation rise with rising F_0 contour. Therefore, we have to use additional information for focus recognition. Moreover, our actual investigations aim at distinguishing 'normal' and 'contrastive/emphatic' focus.

3. THEORETICAL BACKGROUND

Cruttenden [4] defines a kind of 'normal focus' when the focus is located directly before a phrase boundary and is placed on a content word (see example (1b)). Apart from some exceptions focus in another part of a phrase denotes an emphatic or contrastive function. In Ladd [5] we find a detailed description of the problems between syntactic and semantic theories concerning 'normal' vs. 'contrastive' accents. In our case we will only consider the acoustic features of the focus accents and we will neither look at context problems.

maximum - phrase boundary (seconds)	number of CE	number of FA	percentage CE/FA
0.0 - 0.4	10	112	9
0.4 - 0.8	11	75	15
0.8 - 1.2	9	39	23
1.2 - 1.8	10	27	37
Total	43	276	16

Table 1: Measured distances (in seconds) and distribution of contrastive/emphatic (CE) accents vs. distribution of all focus accents (FA)

Examination of our data (see **Table 1**) showed that 40 % of the focus accents are very close (between 0 and 400 ms distance) to a phrase boundary, 28 % are in vicinity (between 400 and 800 ms distance) to a phrase boundary. Considering contrastive/emphatic accents (only 16 % of our focus accents are of this kind), only 23 % of them were found close to a phrase boundary, 25 % were in vicinity. Besides, we found that many contrastive/emphatic focus accents are characterized by an extremely sharp F0 maximum peak while most 'normal' focal accents have a much flatter peak (see for example **Figure 1**).

4. EXPERIMENTS

4.1. Data

The speech material consists of dialogues of German spontaneous speech, supplied within the research project VERBMOBIL. It contains meeting arrangements. Focus accents and contrastive/emphatic accents were labelled for 11 dialogues (195 turns with one or more phrases, 276 focus accents with 43 contrastive/emphatic accents) with 10 different speakers (3 female, 7 male) through acoustic perception. The size of the focus areas was restricted to a word.

4.2. Results

We wanted to examine the correlation between focus accents and phrase boundaries and the relation between 'normal' focus accents and 'contrastive/emphatic' accents respectively. The nearest maximum from every focus accent was computed. The distance between this maximum and the next phrase boundary (in time direction) was measured and this was taken as measure of comparison. Only absolute distances in time were measured, speech tempo was no distinctive factor in our data.

In **Table 1** we see that the percentage of the contrastive/emphatic accents increases with distance from a phrase boundary. In the next measurement distances between the calculated maximum and the next minima to the left and right were computed for both kinds of focus accents.

For all types of focus accents distances between surrounding minima and the maximum are rather similar for all phrase boundary distances (see **Table 2**). There is a small decrease for accents in more than 1.2 seconds distance of a phrase boundary for the left minimum though. On the other hand, the standard deviation for the left minima distances is rather high so that this point is insecure.

maximum - phrase boundary	max - left min	std dev.	max - right min	std dev.
0.0 - 0.4	0.421	0.191	0.337	0.071
0.4 - 0.8	0.478	0.390	0.365	0.092
0.8 - 1.2	0.609	0.590	0.417	0.093
1.2 - 1.8	0.381	0.201	0.354	0.056
1.8 - 3.5	0.495	0.216	0.598	0.252
Total	0.465	0.309	0.379	0.098

Table 2: Measured distances (in seconds) for all focus accents

maximum phrase boundary	max - left min	std dev.	max - right min	std dev.
0.0 - 0.4	0.622	0.324	0.272	0.036
0.4 - 0.8	0.513	0.321	0.442	0.059
0.8 - 1.2	0.495	0.263	0.551	0.133
1.2 - 1.8	0.221	0.065	0.518	0.082
Total	0.482	0.306	0.442	0.082

Table 3: Measured distances (in seconds) for contrastive/emphatic accents

Looking at **Table 3** we find a decreasing left minimum - maximum distance with increasing distance between maximum and phrase boundary. This is especially apparent for contrastive/emphatic accents which have a distance between 1.2 and 1.8 seconds to the next phrase boundary. Moreover, we have a very low standard deviation for the accents in this distance. That could mean that an emphatic accent which is not close to a phrase boundary has a much steeper rise in fundamental frequency compared to the other accents which are close to a phrase boundary.

In further experiments we wanted to examine the range of fundamental frequency, too. The relative distances between maximum and surrounding minima, in respect to F_0 were measured. No correlation for the distance between maximum and phrase boundary and maximum and minimum heights was found.

5. GENERAL DISCUSSION

By integrating information about phrase boundaries from another VERBMOBIL recognition module [6] into the focus recognition algorithm, we have additional help to classify focus accents in 'normal' and 'contrastive/emphatic'. With increasing distance from a phrase boundary there is a higher probability to detect a contrastive/emphatic accent. Moreover, by defining a threshold for 'fast rise' and 'slow rise' in fundamental frequency we have another classification feature. This classification works for at most 55 % of our contrastive/emphatic accents, depending also on the recognition rate for the phrase boundaries (momentarily 81 %). As an additional result we found that by using phrase boundaries, several for the present not detectable focal accents can be found.

Further experiments will try to verify these results with more data. Our data contain obviously very few contrastive/emphatic accents so that it is problematic to generalize the results. Unfortunately it is necessary to label very high amounts of data to find a sufficient number of emphatic accents. And we are still left to manual labelling which is very time consuming.

Moreover it is sometimes desirable to have more 'controlled data' - but this implies a loss in spontaneity. In a sophisticated experimental condition it would perhaps be possible to elicit 'quasi-spontaneous speech' so that we get minimal pairs, i. e. sentences with the same segmental information but with the focus accents on different positions and with different degrees of emphasis.

6. ACKNOWLEDGEMENT

This work was funded by the German Federal Ministry of Education, Science, Research and Technology (BMBF) in the framework of the Verbmobil Project under Grant 01 IV 101 G. The responsibility for the contents of this study lies with the author.

7. REFERENCES

[1] Petzold A. (1995): Strategies for focal accent detection in spontaneous speech. In Proc. 13th ICPhS Stockholm, vol. 3, 672-675

[2] Bruce G., Touati P. (1990): On the Analysis of Prosody in Spontaneous Dialogue, Working Papers, Lund University 36, 37 - 55

[3] Petzold A. (1994): Nachverarbeitung bei der Grundfrequenzbestimmung von Sprachsignalen zur Erfassung von Intonationskonturen, Fortschritte der Akustik - DAGA '94, 1345 - 1348

[4] Cruttenden A. (1986): Intonation. Cambridge University Press

[5] Ladd R. (1978): The structure of intonational meaning. Indiana University Press

[6] Strom V. (1995): Detection of accents, phrase boundaries, and sentence modality in German with prosodic features. Proc. EUROSPEECH'95, Madrid, Spain, 18-21 September 1995, 2039-2041