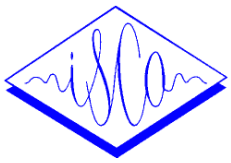

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

ANALYSIS OF VOICE FUNDAMENTAL FREQUENCY CONTOURS OF GERMAN UTTERANCES USING A QUANTITATIVE MODEL

Hansjörg Mixdorff and Hiroya Fujisaki

Department of Applied Electronics, Science University of Tokyo
2641 Yamazaki, Noda, 278 Japan

ABSTRACT

In German, as in many other languages, the fundamental frequency contour (henceforth the F_0 contour) is an important acoustic correlate of intonation. The present paper adopts a quantitative model originally developed for Japanese to analyze German declarative sentences with statement intonation. As far as the current data is concerned the analysis proves that the model is directly applicable to German. The position of the accent command assigned to an accentable constituent depends on whether the item exhibits a terminal or non-terminal accentuation. Phrasing is found to occur at the boundary between clauses but also after larger noun phrases, for instance.

1. INTRODUCTION

Investigations into the prosodic features of a language and their relationship with the underlying linguistic and para-linguistic information are necessary to improve speech analysis and synthesis as well as foreign language teaching [1]. In the case of German, compared with English for example, research activities in this field have been relatively limited. This was widely recognized during the '80s and led to the start of various new projects [2].

An early work [3] by Isačenko and Schädlich showed the importance of pitch and its physical correlate, the F_0 contour, as a prosodic feature of German. In the present paper we will use the term 'intonation' for the prosodic feature expressed by the F_0 contour, being aware of the fact that duration and intensity also play important roles.

The aim of our research is to obtain a precise formulation of German intonation and to find out how it is related to the underlying linguistic units and structures.

2. GERMAN INTONATION AND THE APPROACH ADOPTED IN THE PRESENT STUDY

German has a wide variety of local dialects differing in segmental and supra-segmental features. The present study deals with standard German ('Hochdeutsch') which is used in the nationwide TV news and also promoted as standard for teaching German to foreigners. Although German had traditionally been considered a stress accent language, a recent study [4] showed an interaction of F_0 movement, duration and intensity in German prosody.

In the present study we first examine the F_0 contour of an isolated word uttered with its lexical accent, and see what kind of modification takes place when the same word is being embedded in a sentence. We then present some basic findings about the phrasing of sentences. Finally we deal with the influence of contrastive focusing on the F_0 contour.

In order to gain an insight into the relationship between the F_0 contour and the underlying linguistic units of an utterance we apply the quantitative model by Fujisaki [5], which has originally been developed for Japanese and since been extended to other languages. The original model for Japanese produces an arbitrary F_0 contour by superimposing global (phrase) and local (accent) components. Hence there are two kinds of input signals to the system: impulses (phrase commands) and stepwise functions (accent commands). These are derived in an Analysis-by-Synthesis of the natural F_0 contour. We will try to relate the values derived for these discrete input functions to linguistic units. Figure 1 shows a block diagram of the model and the corresponding mathematical formulations are shown below where $Gp(t)$ denotes the impulse response of the phrase control mechanism and $Ga(t)$ denotes the step response of the accent control mechanism. Fb is the asymptotic value of F_0 in the absence of accent commands.

$$\ln F_0(t) = \ln Fb + \sum_{i=1}^{I} Ap_i Gp(t - T_{0i}) + \sum_{j=1}^{J} Aa_j [Ga(t - T_{1j}) - Ga(t - T_{2j})].$$

$$Gp(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

$$Ga(t) = \begin{cases} \min [1 - (1 + \beta t) \exp(-\beta t), \gamma], & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

The model has already been applied to German [6], but the approach has been significantly different from the one presented in this paper.

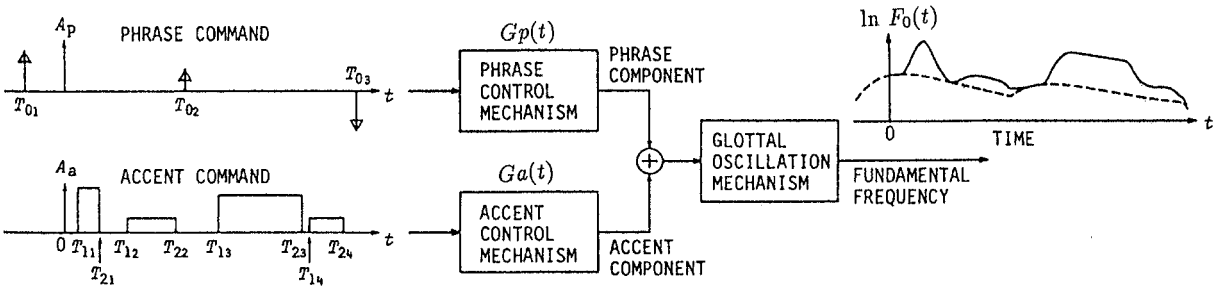

Fig. 1. Quantitative intonation model.

3. SPEECH MATERIAL AND METHOD OF ANALYSIS

The speech material consists of three sets of utterances. Set A contains the target word "Wagen" ("car") uttered in isolation using its citation form.

Set B consists of sentences where the target word is a part of a noun phrase which is successively expanded by adding adjectives before the target word: "Der (neue (helle)) Wagen war an der Wiese." This translates into "The (new (bright-coloured)) car was by the meadow."

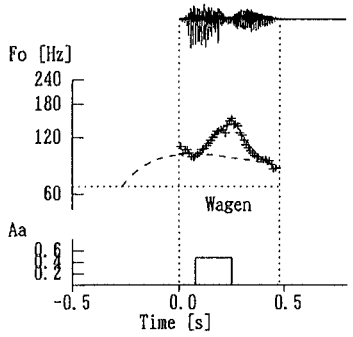

Fig. 2. The target word "Wagen" uttered in isolation.

No focus was specified. Set C contains utterances of the sentence "Sie haben den Wagen geliehen."— "They rented the car," where the focus is placed on various constituents. There is one sentence in set C where prominence is actually placed on a second clause added ("Sie haben den Wagen geliehen und sind TATSÄCHLICH gefahren" — "They rented the car and ACTUALLY drove away.")

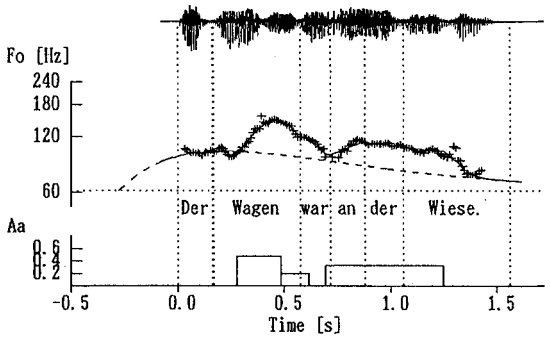

Fig. 3. Example of analysis, "Der Wagen war an der Wiese" by speaker MV.

Four speakers from the northern part of Germany read the sentences at a medium speech rate. In the case of set C the words to be focused were hinted at by embedding the sentences in an appropriate discourse context, while the utterances in sets A and B were produced in isolation. The sentences contain maximally voiced sounds in order to produce continuous F_0 contours. Although the speakers were asked to use 'Hochdeutsch' some of them showed slight dialectal bias.

The utterances were recorded on a DAT and converted at 10kHz (16 bit). After editing of the resulting sound files and marking of word boundaries the F_0 contour was determined using an autocorrelation-based pitch extractor. Possible errors were corrected by listening and visual inspection. The F_0 contour was then modeled in the Analysis-by-Synthesis approach using a graphic editing tool. In the procedure the initial positions and amplitudes of phrase commands are selected by approximately fitting the phrase component to the local minima (the baseline) of the F_0 contour. The accent commands are then placed at the positions of local prominence by first assigning one command to every accentable syllable. The parameter values are then optimized by an iterative procedure for minimizing the mean square error in the ln F_0 domain. Adjoining accent commands with quite similar amplitudes are later merged.

The analysis shows that α and β can be set to constant values ($\alpha = 2.0$, $\beta = 20.0$) without substantially affecting the closeness of approximation.

4. RESULTS OF ANALYSIS

We first examine how the word accent is affected by the function of a target word in a particular utterance.

Figure 2 shows the utterance "Wagen" by speaker MV. At the top of the figure, the speech waveform is displayed. The curve drawn using + symbols indicates the measured F_0 contour, the solid line the synthesized F_0 contour and the dashed line its phrase component part. The accent commands are displayed at the bottom. The slight overall declination of the F_0 contour suggests the presence of a phrase command at 0.28 s before the segmental onset of the utterance. The target

word "Wagen," which has its lexical accent on the first syllable, receives a high accent command.

In Fig. 3 we see the declarative sentence "Der Wagen war an der Wiese" uttered in a neutral intonation by the same speaker. According to rules of German sentence intonation [6], a statement is marked by a fall towards the end of the utterance which generally occurs on the last accentable consituent. In the case of Fig. 3 it is the noun "Wiese" belonging to the adverbial phrase "an der Wiese." The fall coincides with the offset of the accent command assigned to this adverbial phrase. Since 'Wagen' is not the last accented item in the utterance it receives a non-terminal accent pattern. By comparing the portion of the F_0 contour in Fig. 3 belonging to "Wagen" with that of Fig. 2 we can state that in Fig. 3 it does not drop to a low level but stays at a medium level. This can be interpreted as a result of the delayed offset of the corresponding accent command. The accent command in Fig. 3 also exhibits a slightly delayed onset compared to that in Fig. 2. Figures 4 and 5 show utterances where one and two adjectives are inserted before "Wagen." The F_0 pattern belonging to "Wagen" remains quite similar to that in Fig. 3 although the accent command amplitude takes a slightly lower value. This could be due to the fact that prominence is shifted to the adjectives "neue helle." Whereas in Fig. 3 the baseline of the whole utterance can be modeled using

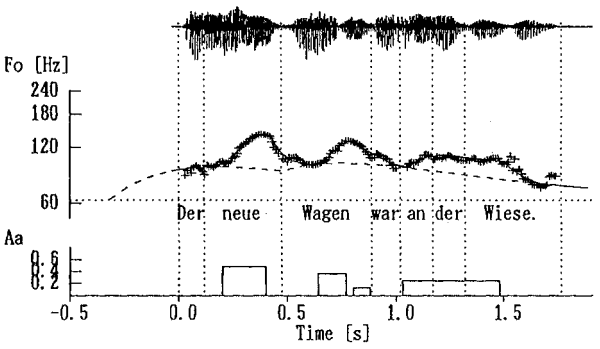

Fig. 4. Example of analysis, "Der neue Wagen war an der Wiese" by speaker MV.

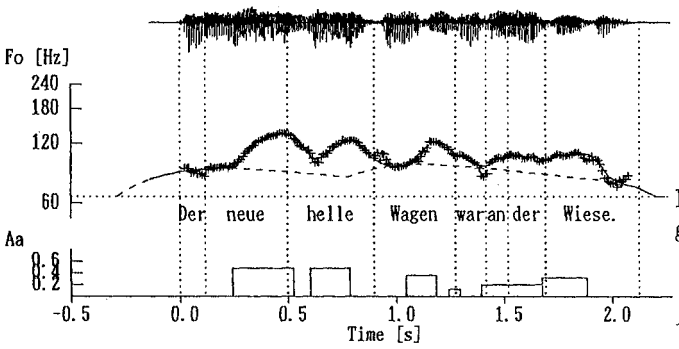

Fig. 5. Example of analysis, "Der neue helle Wagen war an der Wiese" by speaker MV.

Table 1. Parameter values for Figs. 2–5.

	Fig. 2	Fig. 3	Fig. 4	Fig. 5
F_b (Hz)	65.1	62.3	62.4	62.1
A_p	0.4	0.64	0.56, 0.28	0.52, 0.40
Timing of accent comm. for "Wagen"	early	late	late	late
A_a	0.53	0.52	0.36	0.40

only one phrase command, a second phrase command must be added in Figs. 4 and 5 at about 0.45 s before the beginning of the verb phrase "war an der Wiese." In the case of Fig. 5 the second phrase command has a slightly higher level than in Fig. 4. Table 1 contains the model parameters for Figs. 2 to 5.

Figures 6, 7 and 8 show examples of the utterance "Sie haben den Wagen geliehen" by speaker RS. In Fig. 6 narrow contrastive focusing was set on "Wagen" — "They have rented the CAR (not the boat)."

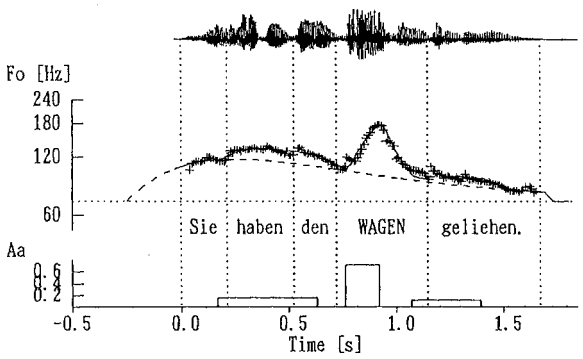

Fig. 6. Example of analysis, "Sie haben den WAGEN geliehen" by speaker RS.

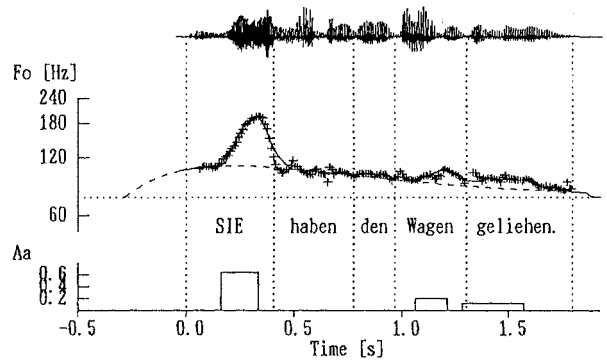

Fig. 7. Example of analysis, "SIE haben den Wagen geliehen" by speaker RS.

The target word "Wagen" exhibits a clearly marked F_0 peak which corresponds to its accent command. The F_0 pattern for "Wagen" is very similar to the one found when the word is uttered in isolation as in Fig. 2. In Fig. 7 "Sie" receives narrow focusing — "THEY rented the car (not their friend)." The portion of the F_0 pattern

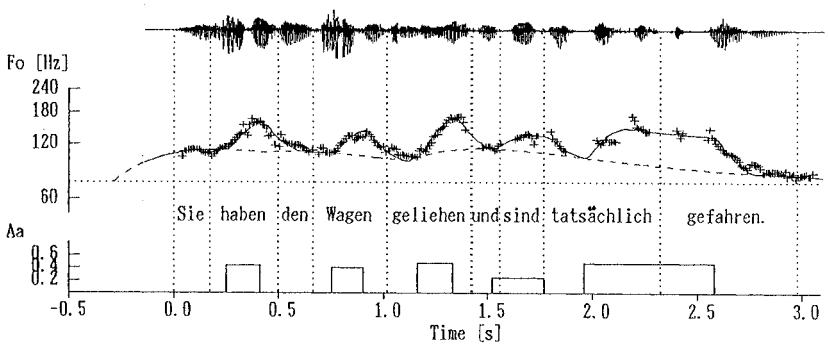

Fig. 8. Example of analysis, "Sie haben den Wagen geliehen und sind tatsächlich gefahren" by speaker RS.

Table 2. Parameter values for Figs. 6–8.

	Fig. 6	Fig. 7	Fig. 8
F_b (Hz)	72.7	74.1	73.7
A_p	0.72	0.48	0.60, 0.44
Timing of accent comm. for "Wagen"	early	late	late
A_a	0.73	0.20	0.40

after "Sie" is almost flat. "Wagen" receives a very low accent command. In Figures 6 and 7 we see once again that the statement intonation is marked by a fall which here occurs on the focused constituent, well before the end of the utterance. Hence we call this accent pattern a 'terminal' accent pattern.

Figure 8 shows the sentence "Sie haben den Wagen geliehen und sind tatsächlich gefahren," where a second clause has been added. Examples 6 and 7 were modeled using a single phrase command. In Fig. 8 an additional phrase command was inserted at about 0.4 s before the boundary between the first and the second clause. The constituents "haben," "Wagen" and "geliehen" receive a non-terminal accentuation and the main prominence of the whole utterance lies on the last two words "tatsächlich gefahren" ("actually drove away !") of the second clause which exhibit a high accent amplitude. Here again the statement intonation is marked by a distinctive fall on "gefahren." Table 2 displays the parameter values corresponding to Figs. 6 to 8. All modeled F_0 contours fit the original ones very well.

5. DISCUSSION AND CONCLUSION

As far as the data presented is concerned, the quantitative model is directly applicable to German statement intonation. Comparison with Japanese data reveals that a single prosodic phrase in German may extend over many more constituents than in Japanese, especially when contrastive focus is present. Phrasing can occur at the boundary between clauses and also at the boundary between a longer noun phrase and the following verb phrase, for example. The amplitude of the phrase command added generally depends on the length of the preceding noun phrase. We found that a single constituent may show a non-terminal (late on- and offset of an accent command) or terminal (early on- and offset of an accent command) accent pattern. The terminal accentuation occurs at contrastively focused items as well as when a target word is uttered using its citation form. In neutral statement intonation, accentable sentence-non-final items receive a non-terminal accentuation.

The prominence given to the individual constituent corresponds to its accent command amplitude. If no focus is specified adjectives may show considerably high amplitudes as compared with the noun which they modify although not as high as in the case of contrastively focused items. Unaccented items are grouped together and show low amplitudes of accent commands.

Our results encourage using the model in the analysis of more complicated utterances. Since the model represents a parsimonious description of the F_0 contour it offers an appropriate basis for the development of a synthesis scheme for rule-generated F_0 contours to be used in a text-to-speech system. This will be one of the main targets of our future work.

REFERENCES

[1] Fujisaki, H. (1993): From information to intonation. *Keynote Lecture at the International Conference on Signal Processing '93, Beijing.*

[2] Altmann, H., Batliner, A. et al. (1989): *Zur Intonation von Modus und Fokus im Deutschen* (Niemeyer, Tübingen).

[3] Isačenko, A.V., Schädlich, H.-J. (1966): Untersuchungen über die deutsche Satzintonation. In *Untersuchungen über Akzent und Intonation im Deutschen* (Akademie-Verlag, Berlin), pp.7-67.

[4] Rusch, M. (1991): *Zur Untersuchung prosodischer Merkmale im Sprachsignal anhand der Sprachgrundfrequenz und der Lautdauer.* Doctor's dissertation at Technical University of Berlin.

[5] Fujisaki, H., Hirose, K. (1984): Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan (E)*, 5, pp. 233-242.

[6] Möbius, B. (1993): *Ein quantitatives Modell der deutschen Intonation* (Niemeyer, Tübingen).

[7] Stock, E. et al. (1982): *Deutsche Satzintonation* (VEB Verlag Enzyklopädie, Leipzig).