# PRODUCTION AND PERCEPTION OF STATEMENT, QUESTION AND NON-TERMINAL INTONATION IN GERMAN

*Hansjörg Mixdorff and Hiroya Fujisaki*

Department of Applied Electronics, Science University of Tokyo

## ABSTRACT

The present study adopts a quantitative model originally developed for Japanese to analyze and resynthesize $F_0$ contours of German utterances. In order to test the descriptive capacity of the model, a production experiment dealing with the realization of statement, question and non-terminal intonation was performed. Based on the analysis results of this experiment, synthetic stimuli for a perception experiment were produced. We examined which of the model parameters are most important for the classification of the sentence mode and found that the accent command offset time T2 largely determines whether an utterance is perceived as statement or unfinished, whereas unfinished and question intonation are distinguished by the accent amplitude.

## 1. INTRODUCTION

Research on the prosodic features of a language and their relationship with the underlying linguistic and paralinguistic information is important to improve speech analysis and synthesis as well as foreign language teaching [1]. An early experiment [2] by Isačenko and Schädlich showed that in the case of German most syntactic and semantic functions can be realized by manipulating the fundamental frequency ($F_0$) contour. 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.

## 2. THE APPROACH ADOPTED IN THE PRESENT STUDY

During recent years much effort has been made to describe the features of German intonation, and to formulate "prototypal" patterns for various sentence types and structures [3].

The present study applies the model by Fujisaki [4] in order to produce a quantitative description of the $F_0$ contour. The model has originally been developed for Japanese and has since been extended to other languages. It 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 by fitting the synthetic $F_0$ contour to the natural one.

We will try to relate the values derived for these input functions to linguistic units. Figure 1 shows a block diagram of the model.

## 3. SPEECH MATERIAL AND METHOD OF ANALYSIS

The speech material consists of utterances of the short sentence with declarative word order "Sie haben den Wagen geliehen."— "They rented the car". The sentence was uttered either in statement or question intonation with a narrow focus on one of the constituents 'sie', 'Wagen' or 'geliehen'. We selected an additional sentence 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.") to examine the realization of non-terminal intonation. The expression 'non-terminal intonation' needs some explanation.

The final intonation pattern chosen for the first clause of a statement consisting of two clauses connected by 'und' depends on the degree of relatedness between clauses [5]. In the unmarked case where the contents of Clause 2 is based in some way on the contents of Clause 1 the $F_0$ pattern exhibits a rise followed by a plateau, whereas in the case of two independent clauses only indicating a sequence in time, both clauses may exhibit a fall at the end like in statement intonation. In our experiment, however, we only deal with the unmarked case which we henceforth mean by 'non-terminal intonation'.



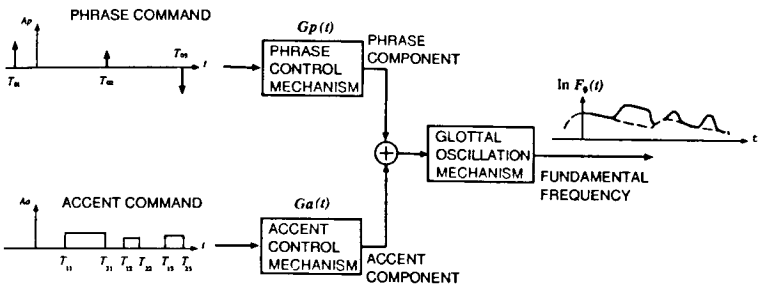


Fig. 1. *Quantitative intonation model.*

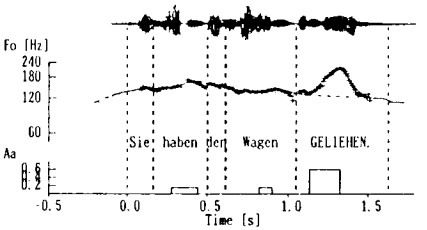


Fig. 2. *Example of analysis, "Sie haben den Wagen GELIEHEN." .*

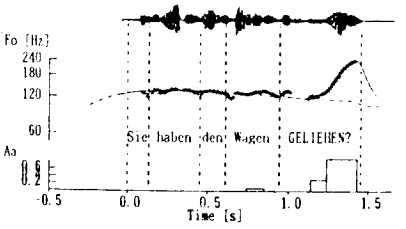


Fig. 3. *Example of analysis, "Sie haben den Wagen GELIEHEN ?".*

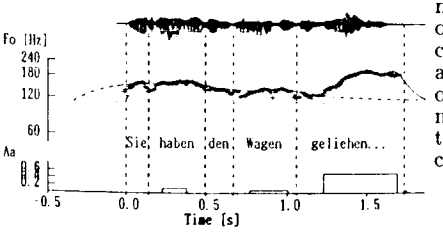


Fig. 4. *Example of analysis, "Sie haben den Wagen geliehen...".*

Twelve speakers from the northern part of Germany read the sentences at a medium speech rate. Words to be focused were hinted at by embedding the sentences in an appropriate discourse context. The utterances were recorded on a DAT and converted at 10kHz (16 bit). After editing the resulting sound files and marking the word boundaries, the $F_0$ contour was extracted. 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 along local minima (the baseline) of the $F_0$ contour. The accent command on-and offsets are mostly aligned with major transitions of the $F_0$ contour ("tone-switches") connected to syllables bearing the word accents. The parameter values are then optimized by an iterative procedure for minimizing the mean square error in the ln $F_0$ domain with $\alpha$ and $\beta$ set to constant values ($\alpha = 2.0$, $\beta = 20.0$).

## 4. RESULTS OF ANALYSIS

In this paper we only discuss the results of the production experiment which are relevant to the perception experiment. We have chosen the condition where a narrow focus is placed on "geliehen". Figures 2, 3 and 4 display examples of analysis for statement, question and non-terminal intonation. At the top of all figures, the speech waveform is displayed. The

**Table 1.** *Mean and standard deviation of accent command timing and amplitude.*
T1: Accent command onset time
T2: Accent command offset time
$A_a$: Accent command amplitude.

| | statem. | question | non-ter. |
|---|---|---|---|
| T1 [ms] $\mu$ | 88 | 170/290 | 220 |
| $\sigma$ | 30 | 50/50 | 50 |
| T2 [ms] $\mu$ | 220 | 290/480 | 440 |
| $\sigma$ | 20 | 50/40 | 20 |
| $A_a$ $\mu$ | 0.55 | 0.45/0.95 | 0.44 |
| $\sigma$ | 0.16 | 0.13/0.13 | 0.06 |

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 statement condition in Figure 2 can be described by a single accent command which is assigned to "geliehen", causing a rise-fall movement of the $F_0$ contour. In question intonation (Figure 3) typically a lower accent command followed by a high one can be observed. The former has a delayed onset compared with statement intonation. Third, the non-terminal intonation (Figure 4) is characterized by a single accent command with on- and offset timing delayed compared with statement intonation. Table 1 gives mean values and standard deviations of accent command timing and amplitudes for the three conditions. The timing is normalized to the mean word duration for all tokens.

It is easily seen that $A_a$ varies considerably within the group of speakers, though auditory check shows no differences in semantic function between the various realizations. The accent command timing is less subject to individual variation.

## 5. THE PERCEPTION EXPERIMENT

Taking the results of the production experiment into account we designed an experiment to examine the relationship between the perception of statement, question and non-terminal intonation and the placement and amplitude of accent commands.

A neutral utterance of the sentence

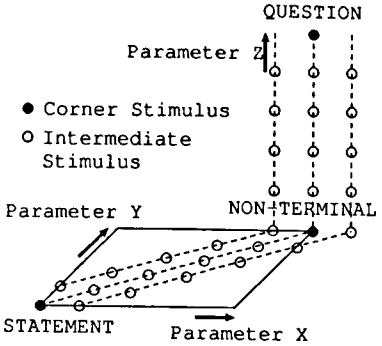


**Fig. 5.** *Locations of synthetic stimuli in the parameter space.*

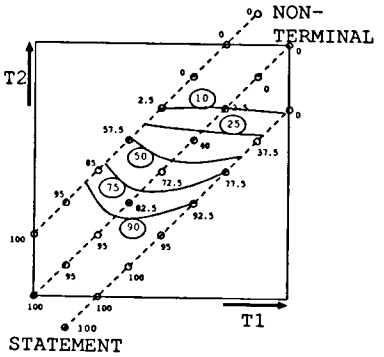


**Fig. 6.** *Probability of judgment 'statement' in the plane T1 vs. T2.*

"Sie haben den Wagen geliehen." is analyzed using LPC of order 14. The data is then resynthesized replacing the original $F_0$ contour by one produced with the Fujisaki model. The phrase component of the synthetic $F_0$ contour is copied from the original $F_0$ contour and remains unchanged for all stimuli. By using the averaged parameter values (see Table 1) for the three conditions (statement, question and non-terminal intonation) we generate three examplary "corner stimuli". These are locations in the multidimensional parameter space as shown for three parameters X, Y and Z in Figure 5. We create intermediate stimuli at equidistant points on the connecting lines between the corner stimuli and additional stimuli along two lines parallel



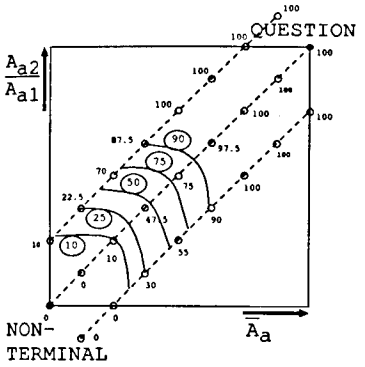


**Fig. 7.** *Probability of judgment 'question' in the plane $\bar{A}_a$ vs. $A_{a2}/A_{a1}$.*

to these. Statement and non-terminal intonation mainly differ as to T1 and T2, whereas the difference between non-terminal and question intonation is characterized by the accent command amplitude ratio $A_{a2}/A_{a1}$ and the amplitude mean value $(A_{a1}+A_{a2})/2$.

20 German subjects (14 male, 6 female), two of whom were trained phoneticians, were exposed to each stimulus five times in random order and were asked to decide, if they perceived it as a statement, a question or an unfinished utterance. They could listen to the stimuli as often as they liked.

## 6. RESULTS

Most subjects consistently identified stimuli grouped around the corner stimuli as belonging to either one of the three cathogories. Figures 6 and 7 show the results of the perception experiment for one pair of corner stimuli each. The locations of the stimuli in the parameter space are marked by dots. In Figure 6, the probability of judgment 'statement' is written to every stimulus, whereas in Figure 7 the probability of judgment 'question' is displayed. By means of maximum likelyhood estimation, lines of equi-probability were determined from the data at 10, 25, 50, 75 and 90 percent levels. The curves suggest that in the case of statement vs. non-terminal the judgment is mainly influenced by the accent command offset time T2. For non-terminal vs. question intonation we find that increasing the accent command amplitude ratio has almost the same effect as increasing the accent command amplitude for both commands.

## 7. DISCUSSION AND CONCLUSION

As far as the data presented is concerned, the quantitative model has proved its applicability to the analysis and synthesis of $F_0$ contours of German. Our perception experiment shows that intonation types can be described by averaged parameter sets. The distinction between statement and non-terminal intonation is mainly determined by the accent command offset time T2. This corresponds to the results of a former study by the authors [6]. Question intonation is characterized by a high accent command amplitude and accent command splitting.

Our results encourage using the model for the formulation of a quantitative model of German intonation and its application to speech synthesis.

## 8. REFERENCES

[1] Fujisaki, H. (1993): From information to intonation. In *Proceedings of the 1993 International Symposium on Spoken Dialogue*, Waseda University, Tokyo., pp. 7-18.

[2] 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.

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

[4] 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.

[5] Pheby, J. (1981): Phonologie: Intonation. In *Grundzüge einer deutschen Grammatik* (Akademie-Verlag, Berlin), p. 890.

[6] Mixdorff, H., Fujisaki, H. (1994): Analysis of Voice Fundamental Frequency Contours of German Utterances Using a Quantitative Model. *Proceedings of the ICSLP '94*, Yokohama, vol. 4, pp. 2231-2234.