PARAMETRIC DESCRIPTION OF GERMAN FUNDAMENTAL FREQUENCY CONTOURS

Bernd Möbius, Grazyna Demenko*, Matthias Pätzold

Inst. f. Kommunikationsf. und Phonetik, Bonn, FRG
*Inst. of Fund. Technol. Research, Poznan, Poland

ABSTRACT

This paper presents a parametric description of German fundamental frequency (F_0) contours as obtained by applying of Fujisaki's intonation model to German. The parameters of the model were extracted by an automatic approximation to naturally produced F_0 courses. The parameter values were standardized using statistical procedures. Finally, intonational prototypes that may be related to linguistic categories were developed by rule. Utterances resynthesized with prototypical F_0 contours were judged highly acceptable by phonetically trained listeners.

1. INTRODUCTION

Intelligibility and naturalness of artificially produced speech may be improved considerably if one allows for prosodic information. Therefore, the generation of prosodic features by rule is an important component of text-to-speech systems. This implies that an adequate description of the intonational variations of the language concerned is at hand. The most outstanding acoustic correlate of intonation is the temporal course of fundamental frequency (F_0).

The aim of this contribution is to separate analytically those factors that determine the F_0 contour of German utterances. This is achieved by applying the quantitative intonation model presented by Fujisaki [1] to German. The model has been elaborated by Fujisaki for the analysis and synthesis of Japanese intonation. It is based on the superposition of a basic value (F_{min}), a phrase and an accent component. The control mechanisms of these components are realized as critically damped second-order systems responding to impulse and rectangular functions, respectively. Thus, the model provides a parametric representation of intonation contours resulting in a considerable data reduction in analysis and synthesis applications.

The extraction of the parameters was attained by a close approximation of naturally produced F_0 curves with the contours generated by the quantitative intonation model. The fitting procedure was implemented in a computer program. Prototypical parameter configurations were derived by statistical analyses and related to linguistic categories, e.g., word accents.

The major issue of this paper is the classification of the parameters and the derivation of intonational prototypes. But first, the speech materials and the method of extracting the parameters will be presented.

2. PROCEDURE

2.1. Speech Materials

In this investigation, the speech material was limited to German declarative sentences containing only one prosodic phrase. 25 test sentences were realized by three male and two female speakers, respectively. For one speaker, the recording was repeated two months

after the first session. The same speaker realized another corpus of 25 test sentences.

The stressed syllables of an utterance were determined in a listening test. By definition, each stressed syllable that is characterized tonally by F_0 movements, will henceforth be called accented. Following Thorsen's [2] description of "stress groups", we define an accent group as a prosodic unit that consists of a leading accent syllable optionally followed by unaccented syllables. This unit is independent of any word boundaries but sensitive to major syntactic boundaries.

2.2. Parameter Extraction

Extraction of the parameter values was done automatically. Using a procedure of analysis-by-resynthesis, the original F_0 course is decomposed into the components of the quantitative intonation model. The parameter values are determined by approximating the contour generated by the model to the original F_0 curve. Based on the principle of superposition, the parameter extraction may be carried out for each component of the model separately.

In our interpretation of the model, the phrase component is considered the baseline of the intonation contour with its maximal value at the very beginning of the utterance. In declarative sentences, a standardized negative phrase command was introduced to allow for the final fall. Each accent group is modelled by the contour resulting from one single accent command. The accent command parameters are determined by the method of least squares.

The parameter values extracted by this procedure were treated statistically with the aim of classification and standardization. The final goal was the derivation of intonational prototypes that are perceptually as acceptable as naturally produced contours. The results of the statistical analysis are presented and discussed in the next section.

3. RESULTS

3.1. Basic Value F_{min}

There is a relatively small dispersion of the basic value F_{min} with all five speakers. 50% of the observed values are found to fall into an interval of about 3.0 Hz around the arithmetic mean of each individual speaker. This small variation makes it reasonable to keep the value of F_{min} constant in experiments with resynthesized speech.

3.2. Damping Factors

The damping factors α and β of the phrase and accent components, respectively, are treated as constants. Fujisaki [1] has shown that the approximation of naturally produced F_0 contours by the model is not impaired if α is assumed to be constant. In our investigation, a fixed value of 3,1 s^{-1} was used. The β range was restricted successively in the present study. Finally, the value of 16,0 s^{-1} proved to be suitable for all speakers and all utterances.

3.3. Phrase Amplitude

Since the phrase component is interpreted as the baseline of the contour, only the phrase amplitude has to be considered here. In an analysis of variance (ANOVA), individual characteristics of the speakers, structure of test sentences and corpora, utterance length, and overall speech tempo were tested as potential sources of variation of the phrase amplitude. The most important factors are *speaker* (F=13.7, p<0.0001) and *sentence* (F=2.9, p<0.001). The significant influence of utterance length was reduced to a strong dependence on the factor *speaker*. Other factors were not found to be significant.

Further analyses revealed that the speakers may be classified into three groups. Taking into account the strong dependence of phrase amplitude on the structure of the sentences, we looked for features common to those sentences with similar phrase characteristics. Since the phrase amplitude is directly related to the steepness of declination, the global downward trend, so our hypothesis, should be stronger especially in those utterances that begin with an accent syllable and end in an unaccented syllable. This hypothesis was confirmed by the result of the ANOVA showing the significant influence of the distribution of accents (F=36.3, p<0.0001).

3.4. Accent Parameters

For this section of the study, a further restriction of the speech materials proved to be necessary. The parameters of the accent component were found to be highly dependent on the position of the respective accent group within the utterance. In order to facilitate the comparison of different utterances, we chose only those sentences that required four accent commands. Thus, the materials consisted of 62 utterances containing 248 accent groups.

With respect to the amplitude of the accent commands, speakers may be grouped into two types. This classification is consistent for all positions of the accent group within the utterance. At the second and third positions, *type of speaker* is the only significant source of variation of the accent amplitude. The duration of the accent group plays a significant role in the initial and in the final position. Furthermore, in utterance initial position, the accent amplitude depends on the word classes: Adjectives require an amplitude value clearly lower than other content words. In the other accent positions, no similar effect of word classes was found.

There is a high positive correlation (r=0.815) between the duration of an accent command and the duration of the accent group to which it is applied. This is also shown by the result of the ANOVA (F=84.1, p<0.0001). No other significant factors determining the duration of accent commands were found.

The temporal distance between the beginning of the accent group and the onset of the command varies with the position of the accent group within the utterance (F=13.7, p<0.0001). While in the first, second and third positions the command is set, on the average, after 10% of the duration of the accent group, the command onset is found immediately after the beginning of the accent group in utterance final position.

Further analysis revealed that there is a direct link between the timing of the accent command and the direction of the accent lending F_0 movement; the effect is significant (F=52.5, p<0.0001). In the speech material under investigation, there is a preponderance of rising and rising-falling movements in the first (98%), second (89%), and third (85%) accent positions. Utterance final accent groups, however, are marked to a large extent (76%) by falling F_0 movements. Since in our interpretation of the model, these movements are approximated by the decreasing part of the contour generated by an accent command, the early command onset in utterance final position is not very surprising. The particular characteristics of final accent commands, differing in some respect from commands in the other positions, is in accordance with the observation that accent groups are sensitive to major syntactic boundaries (cf. section 2.1.).

4. RULE-GENERATION OF F_0 CONTOURS

In the preceding sections, we presented those factors that were found to be responsible for the variation of the parameter values. Standard values were derived subsequently on the basis of the statistically significant factors. A set of rules was formulated in order to generate prototypical intonation contours. By means of LPC analysis and resynthesis, the original F_0 data were replaced by the rule-generated contours.

Acceptability and naturalness of these artificial intonation patterns as well as the adequate realization of the word accents were examined in a listening experiment by six phonetically trained subjects. It turned out that the prototypical intonation contours are highly acceptable: None of the 36 stimuli in the test were rejected by the listeners as being not acceptable or unnatural. With regard to the word accents, more detailed judgements were obtained that led to the improvement of two specific rules, one concerning the slope of the final accent, the other predicting the duration of the accent command in an accent group containing the focus of the utterance. An illustration of a rule-generated intonation contour is given in Fig. 1.

5. CONCLUSION

Fujisaki (1983) has shown that an intonation model based on the superposition principle is a highly useful tool for the analysis and synthesis of the complex F_0 contours in various languages. The effects of different linguistic and speaker-specific features may easily be separated and controlled for, an advantage that facilitates quantitative investigations. The parameters remain constant for a defined stretch of time and may thus be related to linguistic units, in our interpretation to accent groups.

Our application of Fujisaki's work to German has now resulted in a set of rules predicting the parameter values for declarative sentences. Further investigations will have to comprise interrogative sentences, too. The modelling of questions will be particularly interesting, since sentence modality is supposed to be reflected mainly in the phrase component of the intonation model.

ACKNOWLEDGEMENT
This study was supported by the Deutsche Forschungsgemeinschaft and the Alexander von Humboldt-Stiftung.

LITERATUR

[1] FUJISAKI, H. (1983), "Dynamic characteristics of voice fundamental frequency in speech and singing", In P.F. MacNeilage (ed.), *The production of speech*, 39-55, New York: Springer.

[2] THORSEN, N.G. (1989), "Stress group patterns, sentence accents and sentence intonation in Southern Jutland (Sønderborg and Tønder) - with a view to German". *ARIPUC (Copenhagen)*, *23*, 1-85.

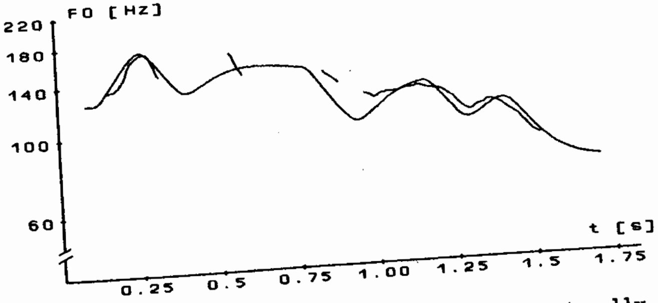

Fig. 1. Rule-generated (continuous line) and naturally produced F_0 contours of the utterance "Hans ißt so gerne Wurst".