Speech Communication 13 (1993) 53–61
North-Holland

Analysis and synthesis of German F_0 contours by means of Fujisaki's model

Bernd Möbius, Matthias Pätzold and Wolfgang Hess

Institut für Kommunikationsforschung und Phonetik, University of Bonn, Poppelsdorfer Allee 47, D-53115 Bonn, Germany

Received 12 May 1993

Abstract. This paper presents the adaptation of Fujisaki's quantitative model to the analysis of German intonation and its application to F_0 synthesis by rule. The parameter values of the model are determined by an automatic approximation of naturally produced F_0 contours. The algorithm is not primarily based on mathematical criteria but is subject to constraints that emerge from a linguistic interpretation of the model. The potential sources of variation of the parameter values are examined using statistical methods. A set of rules is formulated that capture the effects of both linguistic and speaker-dependent features. The rules generate artificial intonation contours which in turn can be related to linguistic features such as sentence mode or word accent. Acceptability of the rule-generated intonation patterns as well as the adequate modelling of linguistic prosodic properties are evaluated perceptually by both phonetically trained subjects and prosodically "naive" listeners. In general, utterances resynthesized with rule-generated F_0 contours are judged highly acceptable and natural by both groups of listeners. Detailed judgements with respect to word accent and sentence mode are obtained that help to improve several specific rules and contribute to a more adequate description of German intonation.

Zusammenfassung. In diesem Beitrag wird die Übertragung des quantitativen Modells von Fujisaki auf die Intonation des Deutschen und seine Anwendung auf die Synthese von F_0-Verläufen nach Regeln vorgestellt. Die Parameterwerte des Modells werden durch automatische Approximation natürlichsprachlich produzierter F_0-Konturen bestimmt. Der Algorithmus basiert nicht in erster Linie auf mathematischen Kriterien, sondern unterliegt Bedingungen, die aus einer linguistischen Interpretation des Modells hervorgehen. Mit Hilfe statistischer Verfahren werden die potentiellen Ursachen für die Variabilität der Parameterwerte untersucht. Es wird ein Satz von Regeln vorgeschlagen, die sowohl linguistische als auch sprecherspezifische Merkmale berücksichtigen. Die Regeln erzeugen künstliche Intonationskonturen, die auf linguistische Merkmale wie Satzmodus oder Wortakzent bezogen werden können. Die Akzeptabilität der regelgenerierten Intonationsmuster wird ebenso wie die adäquate Modellierung von linguistisch-prosodischen Eigenschaften in Perzeptionsexperimenten mit phonetisch ausgebildeten und prosodisch "naiven" Hörern evaluiert. Äußerungen mit regelgenerierten F_0-Konturen werden im allgemeinen von beiden Hörergruppen als sehr akzeptabel und natürlich beurteilt. Durch detaillierte Urteile in Hinsicht auf die Modellierung von Wortakzenten und Satzmodus können einige spezifische Regeln verbessert werden, was zu einer angemesseneren Beschreibung der deutschen Intonation beiträgt.

Résumé. Cette contribution décrit l'adaptation à l'analyse de l'intonation en allemand du modèle quantitatif proposé par Fujisaki et son application à la synthèse de F_0 par règles. Les valeurs paramétriques du modèle sont déterminées par une approximation automatique des contours de F_0 produits en langage naturel. L'algorithme ne s'appuie pas exclusivement sur des critères mathématiques. Il est soumis aux contraintes d'une interprétation linguistique du modèle. Les sources potentielles de la variation des valeurs paramétriques sont examinées à l'aide de méthodes statistiques. Sur la base de ces analyses est formulé un ensemble de règles qui exprime des traits linguistiques aussi bien que relatifs aux locuteurs individuels. Les règles produisent des contours d'intonation artificiels qui peuvent être interprétés relativement à des traits linguistiques, comme par exemple la modalité de phrase ou l'accent du mot. L'acceptabilité des patrons intonatifs produits sur la base des règles ainsi que le modèle adéquat des traits prosodiques sont évalués perceptuellement par des auditeurs professionnels (phonéticiens) et des auditeurs "naïfs". En général, les énoncés resynthétisés avec des contours de F_0 produits par les règles sont jugés très acceptables et naturels par les deux groupes d'auditeurs. Des jugements particuliers sont obtenus sur le modèle des accents du mot et de la modalité de phrase. Ces jugements aident à améliorer quelques règles particulières et contribuent à une description plus adéquate de l'intonation allemande.

Keywords. Prosody; intonation; analysis of F_0 contours; F_0 synthesis by rule.

1. Introduction

Prosodic features of speech signals play an important role in supporting the syntactic and semantic organization of an utterance. Experiments with artificially produced speech have shown that intelligibility and naturalness of the stimuli can be considerably improved if one allows for prosodic information. Therefore, the generation of prosodic features by rule is an important component of text-to-speech systems. In addition, the task of extracting prosodically relevant information from the speech signal also presupposes linguistic and phonetic knowledge about prosodic properties. Both applications imply 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 the fundamental frequency (F_0). Other prosodic features, such as rhythm, tempo, pauses, duration, intensity or voice quality, which are often subsumed under the term intonation in a broader sense, are not in the scope of the investigations presented here, although durations of syllables and stress groups are dealt with implicitly.

The first aim of this contribution is to separate analytically those factors that determine F_0 contours of German utterances. Word accent, prosodic phrasing and sentence mode are considered the most important linguistic factors, while microprosodic effects, such as intrinsic and coarticulatory F_0 variations, are assumed to be irrelevant. The results of this study show that speaker-specific features play an important role, too.

Analysis of German F_0 contours is achieved by applying a slightly modified and extended version of the quantitative model presented by Fujisaki (1983, 1988). This model aims at a functional representation of the production of F_0 contours by a human speaker. Using only a small number of control parameters, the model is able to approximate naturally produced F_0 courses very accurately. For the same reason, this approach promises to be highly useful in perceptual studies of intonation. The decision in favour of Fujisaki's model is motivated in more detail in (Möbius, 1993).

The second major issue of this paper is the classification of the parameter values resulting in a set of rules that generate artificial intonation contours for a given target utterance. The rules capture the effects of both linguistic and speaker-dependent features. Finally, we briefly report on the results of several experiments carried out in order to evaluate the perceptual acceptability of the rule-generated F_0 contours.

2. The model and its linguistic interpretation

2.1. Fujisaki's model

Fujisaki, e.g. (1983, 1988), showed that his quantitative model is a highly useful tool for the analysis and synthesis of complex F_0 contours in various languages. It is based on a hierarchically structured model originally proposed for Swedish by Öhman (1967). The model additively superposes a basic F_0 value ($F_{\min}$), a phrase component, and an accent component on a logarithmic scale (Figure 1). The control mechanisms of the two components are realized as critically damped second-order systems responding to impulse functions in case of the phrase component and rectangular functions in case of the accent component. These functions are generated by two different sets of parameters, i.e., the timing and amplitudes of the phrase commands as well as the damping factors of the phrase control mechanism on the one hand, and the amplitudes and

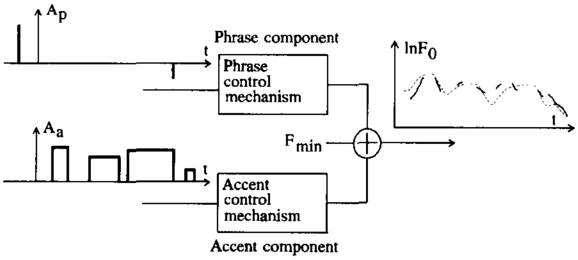

Fig. 1. Block diagram of Fujisaki's quantitative model that additively superposes a basic F_0 value ($F_{\min}$), a phrase component, and an accent component on a logarithmic scale (ln F_0). The control mechanisms of the components respond to impulse (phrase component) and rectangular commands (accent component), respectively (A_p = amplitude of phrase commands; A_a = amplitude of accent commands).

the timing of the onsets and offsets of the accent commands as well as the damping factors of the accent control mechanism on the other hand. All these parameter values are constant for a defined time interval: the parameters of the phrase component for one prosodic phrase, the parameters of the accent component for one accent group, and the basic value F_{min} for the whole utterance.

Using an analysis-by-synthesis procedure, the complex F_0 contour of a given utterance is decomposed into the components of the model. This is achieved by successively optimizing the parameter values which leads to an accurate approximation of the original F_0 course. Thus, the model provides a parametric representation of intonation contours.

So far, the description of the model is in close accordance with the presentations given by Fujisaki (1983, 1988). The model is extremely flexible, its parameters have such a degree of freedom that, in principle, any given contour can be approximated very accurately. But the ability to closely imitate a single F_0 course also raises problems concerning the generalization of the extracted parameter values. It is difficult to relate the configurations of parameter values gained by the approximation of a large number of naturally produced F_0 contours to linguistic categories or features. Fujisaki et al. (1990) give a tentative linguistic foundation of the model by relating prosodic units to syntactic ones but, unfortunately, the authors do not consequently take the next step, i.e., subjecting the approximation itself to linguistic constraints. On the contrary, the approximation of F_0 courses is still based on purely mathematic criteria.

2.2. *An alternative, linguistically based approach*

We propose to proceed the other way round. In our opinion, a quantitative description of intonation is more efficient if a given F_0 contour is structured beforehand into intonational entities that can be more or less directly related to linguistic features. Consequently, modelling naturally produced F_0 courses and extracting the pertinent parameters should be subject to the constraints given by a linguistic and prosodic interpretation in the first place and by the criterion of optimal approximation in a mathematical sense only in the second place.

The phrase component of the model represents the global slope and the slow variations of the F_0 contour in the utterance. Obviously, the phrase component is very suitable to describe F_0 declination since the phrase contour reaches its maximum rather early and descends monotonically along the major part of the utterance. Therefore, the contour which results from adding the basic value F_{min} and the phrase component is interpreted as the baseline of the intonation contour. Apart from the obligatory utterance-initial phrase command, Fujisaki usually provides additional phrase commands, e.g. between a subject and a predicate phrase (Fujisaki et al., 1979). Again, in all the publications referenced, the number of the phrase commands is exclusively determined by the criterion of optimal approximation; additional phrase commands were often inserted even if the utterance had been produced without breath pauses or any other signalling of phrase boundaries.

In German utterances, however, F_0 declination can be described as a feature that spreads over a whole utterance. This observation commonly holds even if an utterance consists of more than one prosodic phrase. In most cases, only major syntactic boundaries, e.g. between main and subordinate clauses, give rise to inserting an additional phrase command that brings about a resetting of the declination line. The conspicuous final lowering of F_0 which is regularly observed in declarative utterances and often in wh-questions can be modelled by a negative phrase command.

The interpretation of the phrase contour as the baseline of intonation also allows us to relate this component of the model to the linguistic category *sentence mode.* As will be shown in Section 3.3, there are both global and local cues on the phrase level that contribute to differentiating between sentence modes.

The local, more rapid changes of F_0 are represented by the accent component of the model. These F_0 movements are superposed onto the global contour and can be related to the realization of accented syllables. To facilitate the linguistic interpretation of the parameter values of the accent commands, we propose to apply an accent group concept: An accent group is defined to be a prosodic unit that consists of an accented syllable optionally followed by unaccented syllables. This is a modification of Thorsen's (1988) well known stress group concept in that accent groups are not delimited by reference to the underlying segmental string but to the F_0 course of the utterance. Accent groups are independent of word boundaries but sensitive to major syntactic boundaries, as will be shown in Section 3.3.

The concept of accent groups is in perfect accordance with the hierarchical structure of the model. While the linguistic category *sentence mode* is reflected in the phrase component, the linguistic feature *word accent* is manifested in the positions and the shapes of the accent commands. Consequently, the F_0 course of a given accent group should be modelled by the contour generated by exactly one accent command. Thus, the parameter configurations of the accent component can be interpreted as correlates of the linguistic feature *word accent*.

The method of determining and standardizing the parameter values of the model will be described in the following section.

3. Analysis of German intonation

3.1. Automatic approximation of F_0 contours

In principle, the parameter values that approximate the F_0 contour of a given utterance can be determined automatically or by hand. Nevertheless, only an automatic procedure guarantees that the optimal values are extracted in an objective and reproducible way. Preliminary experiments showed that there are considerable intra- and interindividual divergencies when an interactive, i.e. partly manual, method is used. Therefore, the parameter values of the model are determined by means of a computer program (Pätzold, 1991) that automatically approximates naturally produced F_0 contours by successively optimizing the parameters within the framework of the linguistic interpretation of the model (cf. Section 2.2). The input information consists of the measured F_0 curve of the given utterance as well as of the accent group boundaries.

Based on the principle of superposition, the step of determining the phrase command parameters and the basic value $F_{\min}$, which is the first step in the algorithm, can be separated from the subsequent determination of the accent command parameters. The contour resulting from $F_{\min}$ and the phrase parameters is approximated to the measured F_0 curve. Once the parameters of the phrase component have been optimized, the resulting difference signal is interpreted by the accent component of the model.

The accent component is made up of partial contours that are in turn generated by accent commands. Each accent group is modelled by the contour resulting from exactly one accent command. The individual accent groups are processed from left to right. There is no global optimization of the whole accent component but a local approximation to the F_0 curve for each accent group. Two restrictions exist that support left-to-right processing of the F_0 contour. The first one prevents that the contour optimized so far is a posteriori affected by a succeeding accent command; the second restriction warrants that the approximation of an accent group is not made impossible by inadequate parameter values of the previous command.

Figure 2 illustrates the close approximation of a naturally produced F_0 contour.

3.2. Speech material

Three different sets of test sentences were used in the present investigation. The first corpus

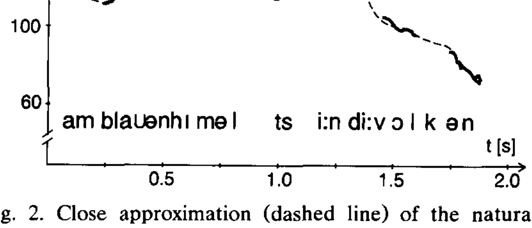

Fig. 2. Close approximation (dashed line) of the naturally produced F_0 contour of the utterance "Am blauen Himmel ziehen die Wolken" (male voice).

is a subset of the so-called "Berliner Sätze" (Sotscheck, 1984). It consists of 25 declarative sentences each containing only one prosodic phrase. This corpus was read by several male and female speakers. Two months after these recordings, one speaker repeated his session with the same test sentences. The same speaker finally produced another subset of 25 test sentences from the "Berliner Sätze".

The second corpus consists of 10 declaratives each containing a main and a subordinate syntactic clause corresponding to two prosodic phrases. The sentences were read by the same speakers. Finally, a third corpus was recorded that consists of 5 test sentences for each of the interrogative modes "wh-question", "yes/no question" and "echo question", respectively. These sentences were produced by one male and one female speaker.

The recordings were made in an anechoic chamber. The speakers were asked to read the orthographically presented sentences aloud. Thus, the speech material used in this study may be characterized as typical "laboratory speech".

3.3. *Statistical analysis of parameter values*

The model parameters extracted from the F_0 contours of our corpora were analyzed using statistical procedures. As potential sources of variation we examined linguistic as well as speaker-dependent factors. The results are presented at full length in (Möbius, 1993), so due to lack of space only the major trends and findings will be discussed here.

3.3.1. *Damping factors*

The damping factors of the phrase and accent components are treated as constants. As it was already shown by Fujisaki (1983), the approximation of naturally produced F_0 contours is not impaired if the damping factor of the phrase component is assumed to be constant. This observation was confirmed by our investigations. As a consequence from the interpretation of the phrase component as a baseline, the exact timing of a phrase command directly depends on the value of the damping factor. With respect to the damping factor of the accent component, the range of values was restricted successively until, finally, a constant value corresponding to the arithmetic mean for all speakers and all accent groups proved to be suitable. Intonation contours generated by the model using constant damping factors were found to be perceptually as similar to the original utterances as were the versions with varying values of the damping factors.

3.3.2. *Basic value F_{min}*

There is a relatively small dispersion of the basic value F_{min} with all speakers. 50% of the observed values were found in the range of about 3.0 Hz around the arithmetic mean F_{min} value for the respective speaker. This very small variation suggests that in experiments with resynthesized speech it will be reasonable to keep F_{min} constant for a given speaker.

Furthermore, there is a negative correlation (ranging from -0.57 to -0.77, depending on the speaker) between the values of the parameter F_{min} and the phrase command amplitudes. The lower the course of F_0 is at the end of the utterance – which means that the declination effect is very salient – the higher the value of the phrase command amplitude has to be. Therefore, the magnitude of the phrase command amplitude can be interpreted as a measure of F_0 declination.

3.3.3. *Phrase command amplitude*

The amplitudes of the phrase commands are largely speaker dependent. Three types of speakers were found differing significantly with regard to the phrase command amplitudes. Another important factor is the distribution of word accents within the utterance. Utterances beginning with an accented syllable and ending with an unaccented syllable exhibit significantly (at $p < 0.001$) higher values of the phrase command amplitude than utterances with any other word accent structure. This result is in no way surprising, since in the speech material under investigation by far the most utterance-final accents of declaratives are marked by falling F_0 movements that result in a steep slope of declination.

Furthermore, the sentence mode is globally signalled by the contour of the phrase component. While the phrase contours of wh-questions are very similar to those of declaratives, yes/no-questions and the syntactically unmarked echo questions show a much less steep declination. Moreover, these types of utterances are globally marked by a positive final phrase command.

In the utterances containing two prosodic phrases we found a high positive correlation between the amplitudes of the phrase commands. The values of both commands tend to be higher if the main clause corresponds to the second prosodic phrase, and lower if it coincides with the first prosodic phrase. Obviously, the position of the main accent or focus in the utterance plays an important role in explaining this finding.

No dependency of phrase command amplitude on utterance duration or speech tempo was observed.

3.3.4. Accent command amplitude

Two types of speakers were classified that differ significantly (at $p < 0.001$) with respect to the amplitudes of the accent commands. Another important factor is the position of the respective accent group in the utterance. Utterance-final accent commands show significantly (at $p < 0.001$) smaller amplitudes than accent commands in any other position in the utterance.

In addition to these most important factors, the type of word carrying the accent may serve to adjust the amplitude value in that nouns seem to require higher values (about 10%) than any other word class. This tendency is significant (at $p < 0.001$).

Furthermore, the amplitude of an accent command seems to be sensitive to sentence mode. In yes/no and echo questions, the final rise of the intonation contour, which is mainly approximated by a positive utterance-final phrase command, is supported by a higher value of the final accent command. This tendency is especially clear in syntactically unmarked echo questions.

In the utterances containing two prosodic phrases, the phrase boundary is not only signalled by an additional phrase command but also by the characteristics of the accent command immediately preceding the boundary. The accent command amplitude in this position tends to be about 25% higher than in a comparable accent group not preceding a phrase boundary. There are speaker-specific discrepancies, however.

3.3.5. Accent command duration

The duration of an accent command can be reliably predicted from the duration of the respective accent group. There is a high correlation ($r = 0.84$) between these two variables, and more than 70% of the variance observed in the durations of accent commands can be explained by the accent group duration ($r^2 = 0.71$).

An effect of phrase-final lengthening is observed for several speakers. In the utterances with two prosodic phrases, accent commands immediately preceding the phrase boundary tend to be considerably longer than those in other positions, again directly depending on the duration of the accent group concerned.

3.3.6. Accent command position

The most important factor controlling the relative temporal position of an accent command within a given accent group is the position of the accent group in the utterance. While in non-final positions the temporal distance between the beginning of the accent group and the onset of the command is about 10% of the accent group duration, this distance tends toward zero in utterance-final accent groups.

4. Generating intonation contours by rule

In the preceding section, the potential sources of variation of the parameter values were explored by means of statistical analysis. Standard values were derived on the basis of the statistically significant factors. A set of rules was formulated that control the adjustment of the parameters (see (Möbius, 1993) for details). The rules capture speaker dependent as well as linguistic features such as sentence mode, sentence accent, phrase boundary signals or word accent, and generate an artificial intonation contour for any given target utterance. The input information needed for generating an intonation contour by rule is the temporal position of accented syllables. Currently, our F_0 synthesis is confined to rather short utterances containing not more than two prosodic phrases. An illustration of an intonation contour generated by rule is given in Figure 3.

The adequacy of the rules was critically examined by expert and "naive" listeners in a series of perceptual experiments described in Sections 4.1 and 4.2. Informal listening tests that preceded the experiments presented here showed that a pair of utterances, one with the original F_0 information and the other with a close approximation of the original intonation contour, could not be discriminated by prosodically "naive" listeners in a pairwise comparison. The perceptual identity of a naturally produced F_0 contour and its close approximation by the model proves that the exponentially damped smooth contours generated by the model evoke a perceptual impression that is as acceptable as that of natural F_0 patterns. Additionally, it is obvious that the most important features of natural intonation contours are preserved in the modelled contour. This is an important aspect since the statistical analysis that aims at a standardization of the parameter values refers to the results of this approximation.

4.1. Acceptability of artificial intonation contours: preliminary results

In a perceptual experiment, we aimed at assessing the validity and correctness and also the potential shortcomings of the rules.

METHOD. The original F_0 data of 25 utterances (5 test sentences from the "Berliner Sätze", each produced by 3 male and 2 female speakers) were replaced by rule-generated contours. Additionally, the test corpus also included 10 utterances with their original intonation contours. The stimuli were judged by six phonetically trained listeners both globally concerning the acceptability and the naturalness of prosodic properties and in more detail regarding the adequate modelling of sentence mode and word accents.

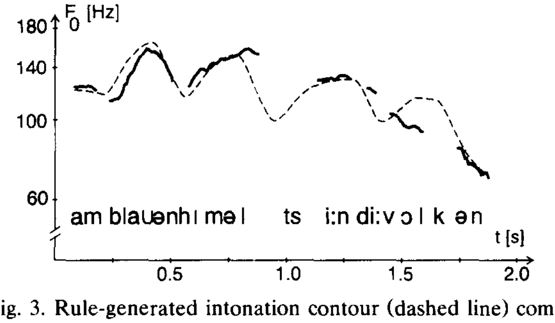

Fig. 3. Rule-generated intonation contour (dashed line) compared to the original F_0 contour of the utterance "Am blauen Himmel ziehen die Wolken" (male voice).

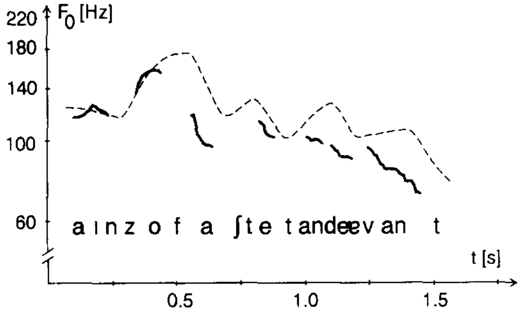

Fig. 4. The rule-generated intonation contour (dashed line) for the declarative utterance "Ein Sofa steht an der Wand" (male voice) with a word accent on the final syllable inadequately evokes the perceptual impression of a non-terminal contour.

RESULTS AND DISCUSSION. The intonation contours generated by rules were found to be highly acceptable. None of the 35 stimuli in the test were rejected by the listeners as being unacceptable or unnatural. With respect to the word accents, more detailed judgements were obtained that concentrated on two problematic aspects. On the one hand, the final descending slope of F_0 is not distinct enough in some of the utterances with a word accent on the final syllable (Figure 4). Perceptually, this shortcoming results in the impression of a non-terminal intonation contour that does not match the intended declarative sentence mode. On the other hand, the third accent command is too long in all the versions of the test sentence "Günter muß noch einkaufen gehen", regardless of the speaker, which results in an accent shift from the syllable "ein-" to the syllable "-kau-" (Figure 5).

MODIFICATION OF RULES. On the basis of the detailed judgements by trained phoneticians, the specific rules concerned were modified. The perceptual impression of a progredient intonation contour was corrected by a reduction of the duration of utterance-final accent commands. The second instance is more complex. Although a temporal reduction of the accent command would be appropriate in the utterance used in the test, there are utterances in our corpus containing

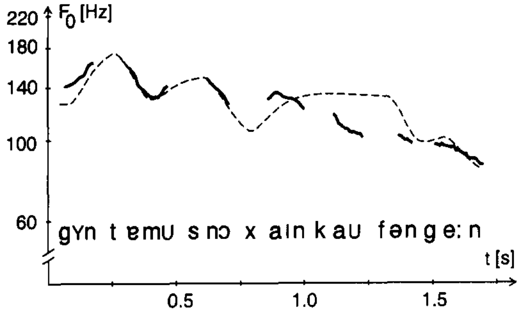

Fig. 5. The inadequate third accent command in the utterance "Günter muß noch einkaufen gehen" (male voice) results in an accent shift from the syllable "ein-" to the syllable "-kau-".

accent groups of comparable structure and duration that do require an accent command just as predicted by the rules. Thus, the factor distinguishing those cases from the one in the test has yet to be determined.

4.2. Perceptual evaluation

Based on the slightly revised version of the rules, three perceptual experiments were carried out in order to evaluate the acceptability of artificial intonation contours as well as the adequate realization of prosodically relevant linguistic features (Möbius and Pätzold, 1992a). Two groups of listeners participated in these experiments, 4 phonetically trained staff members of our institute and 18 "prosodically naive" undergraduate students. The subjects were asked to judge the stimuli by their melodic and stress features and to abstract, as far as possible, from voice and sound quality.

One female and one male "voice" were used in the experimental sessions. There were three kinds of stimuli to be evaluated: (a) naturally produced utterances with their original F_0 information; (b) utterances with close approximations of the original F_0 course generated by the model; (c) utterances with rule-generated intonation contours. The F_0 manipulations for versions (b) and (c) were achieved using the PSOLA technique (Moulines and Charpentier, 1990).

The purpose of the first experiment was to assess the quality of the close approximations as against the respective originals in a pairwise comparison. The listeners were asked to decide by forced choice which of the two versions in each pair was better. Based on earlier observations, our hypothesis was that the subjects would not be able to discriminate between the originals and their close approximations which would result in a chance distribution of the judgements. The results indicate, however, that both groups of listeners were able to tell the two versions of an utterance apart, although presumably mainly on the basis of some unavoidable audible effects of the F_0 manipulation and not by perceived differences in the intonation contours. The listeners were not consistently able to base their discrimination on prosodic information.

In the second experiment, utterances with close approximations of the original F_0 contour as well as utterances with rule-generated intonation contours were presented to the listeners. Subjects expressed their ratings of the prosodic properties of the stimuli on a five-point scale. The rule-generated contours for declaratives were generally not much less acceptable than the close approximations of the respective original contours, while the ratings for rule-generated interrogatives were significantly lower. In general, the experts' ratings were considerably higher than those of the naive subjects.

The third experiment aimed at obtaining detailed judgements with respect to the realization of word accents and sentence mode. Annotations made by the listeners pointed at potential reasons for the results of the previous experiment. We are confident that deaccentuation rules for compound nouns and for adjacent accented syllables as well as the integration of falling–rising F_0 patterns for specific types of interrogatives into our model will contribute to a more adequate description of German intonation.

5. Work in progress

Apart from the analysis of German intonation in a strict sense, the quantitative model is currently applied in two research projects at our institute, one dealing with speech synthesis based on the concatenation of non-parametric units such as demi-syllables, diphones and suffixes (Portele et al., 1990), the other one aiming at the development of a prosodic component within the framework of a joint national project in speech recognition (Pätzold and Möbius, 1991).

The task in the latter project is to extract prosodically relevant events from the speech signal. The characteristics of the F_0 contour are interpreted by means of Fujisaki's model. The procedure described in Section 3.1 has recently been refined and developed towards a stand-alone program without any input information apart from the F_0 curve of the utterance to be analyzed. The detection of the prosodic event "accentuation" is achieved by automatically determining accent group boundaries and the positions of the accent commands. The parameter values are incrementally optimized, and the underlying assumption is that the values tend to converge within a given accent group but change significantly whenever a new command is needed (Möbius and Pätzold, 1992b).

Acknowledgments

This research was supported in part by grants from the German Research Council (DFG) and the German Federal Ministry of Research and Technology (BMFT).

References

H. Fujisaki (1983), "Dynamic characteristics of voice fundamental frequency in speech and singing", in *The Production of Speech*, ed. by P.F. MacNeilage (Springer, New York), pp. 39-55.

H. Fujisaki (1988),"A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour", in *Vocal Physiology: Voice Production, Mechanisms and Functions*, ed. by O. Fujimura (Raven, New York), pp. 347-355.

H. Fujisaki, K. Hirose and K. Ohta (1979), "Acoustic features of the fundamental frequency contours of declarative sentences in Japanese", *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics (Univ. of Tokyo)*, Vol. 13, pp. 163-172.

H. Fujisaki, K. Hirose and N. Takahashi (1990), "Manifestation of linguistic and paralinguistic information in the voice fundamental frequency contours of spoken Japanese", *Proc. Internat. Conf. Spoken Language Processing*, Kobe, Japan, Vol. 1, pp. 485-488 (paper 12.1.1).

B. Möbius (1993), *Ein quantitatives Modell der deutschen Intonation. Analyse und Synthese von Grundfrequenzverläufen* (Niemeyer, Tübingen).

B. Möbius and M. Pätzold (1992a), "F_0 synthesis based on a quantitative model of German intonation", *Proc. Internat. Conf. Spoken Language Processing*, Banff, Alberta, Canada, Vol. 1, pp. 361-364.

B. Möbius and M. Pätzold (1992b), "Bonner Arbeiten zur Intonation", *ASL Prosody Workshop*, Bonn, 23 September 1992, presented.

E. Moulines and F. Charpentier (1990), "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones", *Speech Communication*, Vol. 9, Nos. 5/6, pp. 453-467.

S.E.G. Öhman (1967), "Word and sentence intonation: A quantitative model", Royal Inst. of Technology (Stockholm), *STL-QPSR*, Vols. 2-3, pp. 20-54.

M. Pätzold (1991), Nachbildung von Intonationskonturen mit dem Modell von Fujisaki. Implementierung des Algorithmus und erste Experimente mit ein- und zweiphrasigen Aussagesätzen, Unpublished M.A. Thesis, Univ. of Bonn.

M. Pätzold and B. Möbius (1991), "Ein Konzept für das akustisch-phonetische Prosodie-Modul in ASL-Nord", *Workshop Prosodie in der Mensch-Maschine-Kommunikation*, Erlangen, 2-3 December 1991.

T. Portele, W.F. Sendlmeier and W.J. Hess (1990), "HADIFIX: A system for German speech synthesis by concatenation of non-parametric units", *Proc. ESCA Workshop on Speech Synthesis*, Autrans, France, pp. 161-164.

J. Sotscheck (1984), "Sätze für Sprachgütemessungen und ihre phonologische Anpassung an die deutsche Sprache", *Fortschritte der Akustik - DAGA '84* (DPG, Bad Honnef), pp. 873-876.

N.G. Thorsen (1988), "Standard Danish intonation", *Ann. Rep. Inst. Phonet. Univ. Copenhagen*, Vol. 22, pp. 1-23.