Towards a Model for German Prosody

Robert Bannert

ABSTRACT

In this paper, an attempt is made to outline a model for German prosody. The model aims at generating the rhythm and melody of German utterances, i.e. their temporal and tonal structures, by starting from linguistic information in the input and then applying phonological and phonetic rules. The model consists of three main components: the basic temporal and tonal components where the basic temporal and tonal structures of utterances are generated separately from each other and the modification component where the basic structures are modified in different ways according to various demands. Starting out from an inherent duration of the segments spectrally defined in the input, their basic durations are calculated as a consequence of various contextual factors. Operating on the tonal prosodic features of the input, pitch rewriting rules transpose the linguistic features into a notation of tonal points or levels HIGH or LOW which are related to spectral events and of the tonal range WIDE bound to the HIGHs and LOWs. These tonal specifications are transferred in a series of steps by the intonation algorithm into phonetic values. Drawing on acoustic data and on manipulations of the pitch of utterances by means of speech synthesis, some concepts of an intonation theory concerning German are discussed.

1. BACKGROUND AND AIM

In contrast to what is the case for a great number of languages, there does not seem to exist a model for German prosody which is based on acoustic data and which, although qualitatively only, can generate the temporal and tonal structure of utterances by means of rules[1]. Prosody is used here in a somewhat limited sense; the component of voice quality and certain segmentally bound features are excluded. Thus prosody means simply the rhythm and melody of speech.

Rhythm is equivalent to the temporal structure, namely the durations of segments and their interrelationships within larger units such as syllables, phrases, etc. Melody is equivalent to the tonal structure, expressed, for instance, as the Fo-contour which is composed of tonal prosodic features such as accent, intonation type, emphasis, etc. The tonal structure of utterances may be viewed from different angles. From the point of view of the speaker, the tonal contours may be described as either movements or points or levels representing tonal targets. From the point of view of the listener, the tonal structure may be seen in terms of tones or tonal changes associated with certain spectral events such as vowels, consonants, and syllables.

It is the aim of this paper to present an outline of a model for German prosody[2]. I would like to emphasize that this cannot be more than just a first, simple, and incomplete attempt; a great amount of work and research needs to be done to enable us to achieve a more complete and satisfactory understanding of German prosody.

The present outline is based on an acoustic investigation where several important parameters were varied within utterances spoken as one breath group[3]. The second basis for this sketch is to be found in the various models for duration and intonation which have been developed for some other languages (see references). These models served as helpful guide-lines for the more general aspects of the German prosody model.

2. SOME ACOUSTIC DATA OF GERMAN PROSODY

A relatively large material which was read by three university educated female speakers from Northern Germany served as the acoustic basis for the outline of the model for German prosody. The material which is shown in Table 1 consisted of 14 basic utterances[4] the length of which varied as a consequence of the increasing number of accents from 1 to 8. The main syntactic phrase boundaries are indicated.

All utterances were read by each speaker using the following three intonation types: statement (final falling intonation), echo question (same word order, but expressing surprise and

Table 1. The test sentences containing from 1 to 8 accents. Sentences with 2 to 7 accents appear as pairs differing in the position of the syntactic phrase boundaries which are indicated. The test sentences were read as the intonation types statement (A), echo question (E), and information question (I) with inverted word order.

number of accents	sentence number	utterances
1	1	# Die Männer. #
2	2	Die längeren Männer.
	3	Die Männer in der Menge.
3	4	Die längeren Männer in der Menge.
	5	Der Müller will die Männer immer Lümmel nennen.
4	6	Der Müller will die längeren Männer in der Menge nennen.
	7	Der Müller in Lingen will die Männer immer Lümmel nennen.
5	8	Der lullende Müller in Lingen will die Männer immer Lümmel nennen.
	9	Der Müller will die längeren Männer in der Menge immer Lümmel nennen.
6	10	Der lullende Müller in Lingen will die längeren Männer in der Menge nennen.
	11	Der lullende Müller in Lingen will die längeren Männer immer Lümmel nennen.
7	12	Der lullende Müller will die längeren Männer in der Menge immer lungernde Lümmel nennen.
	13	Der lullende Müller in Lingen will die längeren Männer immer lungernde Lümmel nennen.
8	14	Der lullende Müller in Lingen will die längeren Männer in der Menge immer lungernde Lümmel nennen.
		# #

astonishment, final rising intonation), and information question (inverted word order, final rising intonation). The speakers aimed at giving equal weight to each accent and at producing each utterance as one prosodic phrase without breaking up the longer utterances into several shorter phrases. The utterances containing 2 to 7 accents appeared as pairs differing in the position of the syntactic phrase boundaries. Each utterance was embedded in an appropriate context (cf. Bannert 1983a and for the method Bruce 1977) and read 5 times.

For the sake of convenience, and before outlining the model proper, some basic data concerning certain temporal and tonal aspects of German prosody will be given.

2.1 Sentence and stress group durations

It is well known that segment durations are affected by a number of segmental and prosodic factors. In this respect, the following question is of interest. Are the durations of sentences and stress groups affected by the tonal structures of the different intonation types? This question can only be given a negative answer (for a detailed description cf. Bannert 1983b). The durations of the units of sentence (utterance) and stress group do not vary in a uniform way across the intonation types and the three speakers. No systematic patterns of variation emerge that can be interpreted as a common temporal behaviour reflecting specific linguistic demands shared by all the speakers.

The relationships between the sentence durations of the three intonation types A (statement = reference), E (echo question), and I (information question), according to two methods of evaluation, can be summarized as follows:

$$I \leq A \leq E$$

This means that the three sentence durations are either equal or unequal, i.e. there may or may not be any significant difference between them. If there is a difference, however, the information question has the shortest and the echo question the longest duration. Nevertheless, it should be remembered that the relative differences in sentence duration between the intonation types and across the three speakers are rather

small. They vary in a non-systematic way between 0 and 9% of the duration of the statement. As the differences in sentence duration are not systematic and of minor magnitude only, it may be assumed therefore that the differences observed may be perceptually insignificant.

The durations of the stress groups show a similar picture. They do not vary consistently between the three intonation types and across the three speakers. Variation of stress group durations are to be found over the whole utterance. They are not confined to any specific part of the utterance, e.g. the final part where the tonal contour moves in opposite directions (statement vs. questions). Thus it is inferred that the differences in sentence duration observed in the material are the result of a global temporal change which is distributed over the whole utterance.

By means of a graphic representation, the non-systematic pattern of temporal variation at the sentence and stress group level becomes obvious. Speaker B's data are used which are also representative of the two other speakers. Figure 1a shows the relative mean differences of sentence duration between the intonation types statement/information question and echo question/statement of all the 14 utterances. Figure 1b shows the relative mean differences of stress group duration between echo question/statement and their total mean of the longest utterance with 8 accents for all three speakers.

In summary, then, it is concluded that the differences in sentence and stress group duration between the three intonation types are relatively small and by no means systematic or consistent. Therefore, it seems justified, at least as a first approximation, to attribute no obligatory (linguistic) status to the context feature "intonation type". Instead, the duration of intonation types may vary individually. This variation may therefore be considered optional to be incorporated as such into a later stage of the prosody model.

2.2 Fo-contours

As an illustration of the tonal features and their systematic characteristics, Figure 2a shows the normalized and superimposed Fo-contours of the utterance with 6 accents (Der

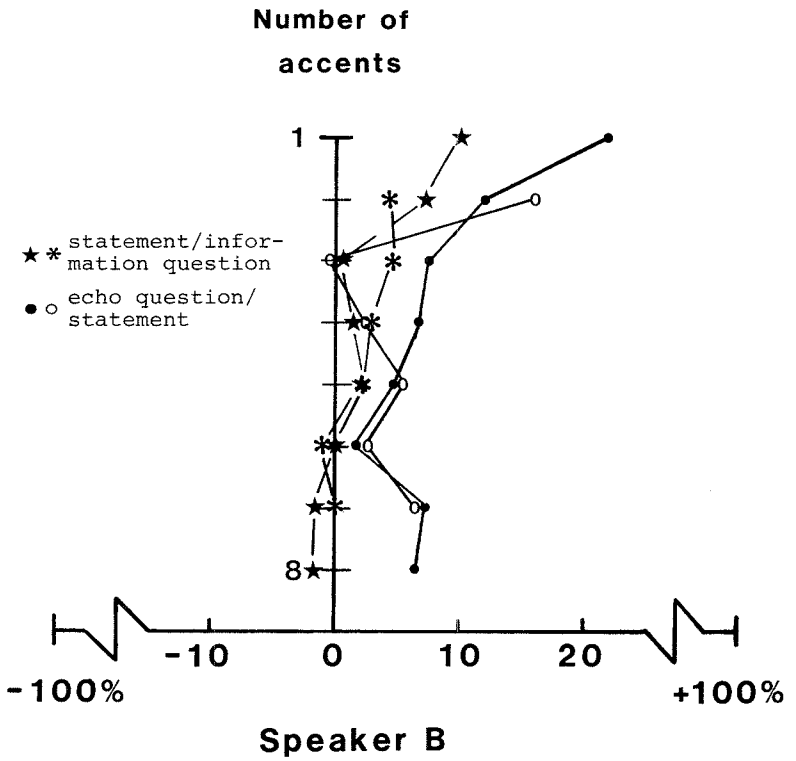


Figure 1a. Relative mean differences of sentence durations between sentence types in the 14 utterances with 1 to 8 accents, statement duration being the reference.

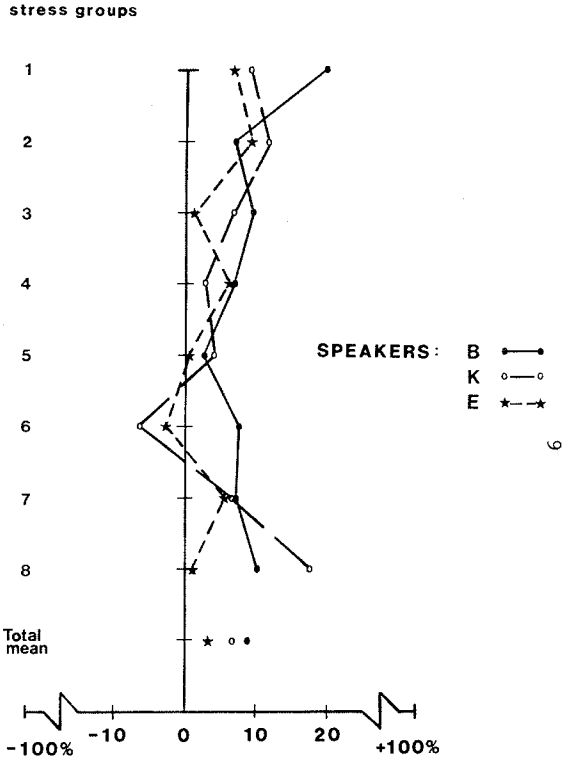


Figure 1b. Relative mean differences of stress group durations in sentence 14. See text.

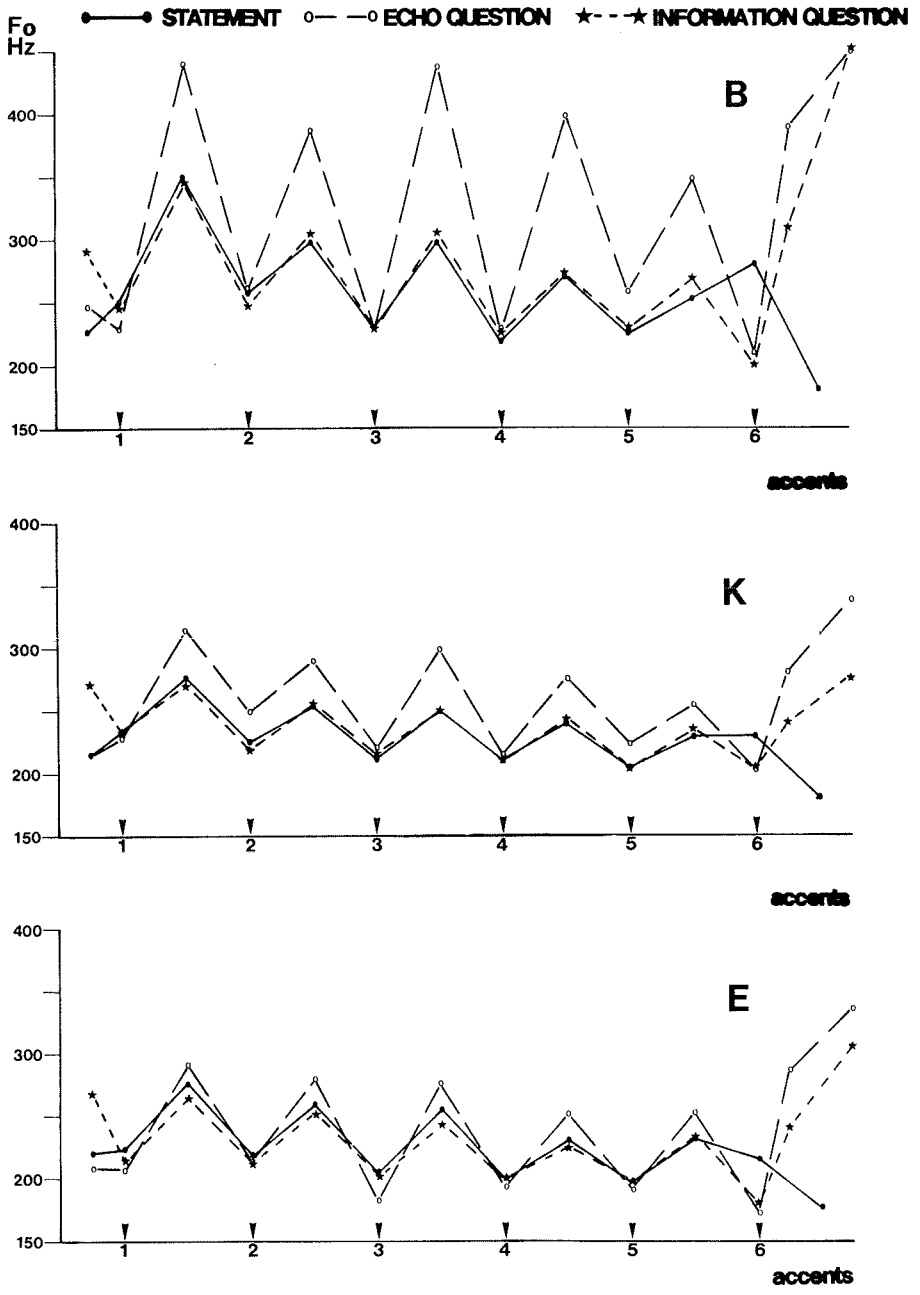


Figure 2a. Normalized and superimposed Fo-contours of the utterance 11 with 6 accents, spoken as a statment, echo question, and information question by three speakers B, K, and E.

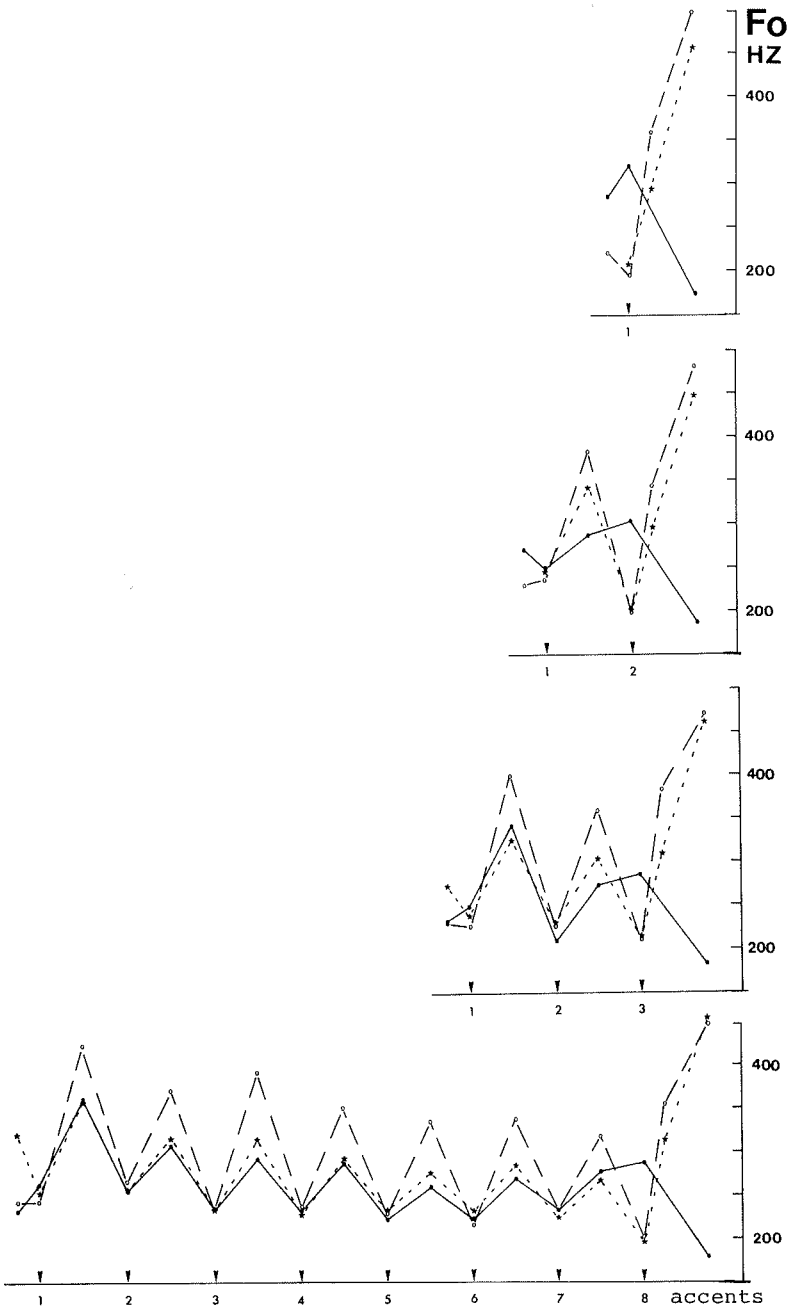


Figure 2b. Normalized and superimposed Fo-contours of four utterances containing 1,2,3,and 8 accents of speaker B. Line-up point is the end of the tonal contour bringing out the systematic structure of the contours.

lullende Müller in Lingen will die längeren Männer immer Lümmel nennen) as the three intonation types for each speaker. Except for one instance, the tonal expression of each accent consists of a tonal rise starting from an Fo-minimum in the consonant preceding the accentuated vowel and ending in an Fo-maximum in the following unstressed syllable. There is only one exception, namely the final accent of statements which is manifested tonally as a fall. It will be recognized that the tonal contour of the two final accents in statements (rising-falling) equals the hat pattern typical for Dutch intonation ('t Hart and Cohen 1973). The tonal differences between the intonation types, above all, are to be found locally at the end of the utterances. The statement is characterized by a rather steep fall in the final accentuated vowel, continuing to the very end of the utterance and ending in a very low tone, the speaker's absolute Fo-minimum. The questions, on the other hand, end in a rising contour reaching a very high tone well above the tonal peaks of the preceding accents. In general, the echo questions show an increased tonal range in the accents, the size of which varies between the speakers. Considering the three superimposed tonal contours representing the three intonation types, one may get the impression that the tonal movements started from a common tonal floor. In most cases, the Fo-minima of all three intonation types differ only slightly. The Fo-maxima, however, may differ greatly, especially for speaker B. In other words, the tonal range between intonation types and speakers varies considerably, the Fo-minima serving as a fixed floor.

It is very easy, starting from Figure 2a, to predict the tonal contours of the other utterances containing 1-5 and 7-8 accents. The Fo-contour of the utterance with 1 accent (Die Männer) by and large corresponds to the contour associated with accent 6, i.e. the final accent, in Figure 2a. The tonal contours of utterances with 2 accents correspond to those of accents 5 and 6; those with 3 accents to those of accents 4, 5, and 6, etc. To be added is the very beginning of the contours, i.e. the initial part preceding the Fo-minimum of the first accent, as shown in Figure 2a (cf. also Bannert 1983a, b).

The systematic nature of the tonal structure when the number of accents is increased, and thus the length of the utterance is also increased, is brought out very clearly in Figure 2b where the normalized and superimposed Fo-contours of utterances with 1, 2, 3, and 8 accents for Speaker B's intonation types are shown. Going from 1 to 8 accents, the final part of the contour, i.e. that contour which is associated with the final accent, is maintained while rising-falling contours, corresponding to each non-final accent, are added before the final accent.

3. THE COMPONENTS OF THE PROSODY MODEL

It is the aim of the prosody model to generate the temporal and tonal structure of a given utterance by means of rules, i.e. by applying phonetic information in a systematic way to a given input structure. In principle, this goal may be achieved in different ways. Two questions, however, should be answered in advance as they have far-reaching implications for the design of such a model.

Firstly, the two-fold aim, namely to generate two structures - one in the dimension of time and the other in the dimension of fundamental frequency, which, in turn, is a function of time -, represents some special difficulties in itself. From the very start, the basic question has to be asked whether the two structures exist each one in its own right, or if one of them can be derived from the other. The answer to this question must have great significance for the structure of the model. If one of the two structures can be derived from the other, then it would not only mean technical savings in the design of the model, but such a relationship, above all, would capture the very nature of the relationship between the temporal and tonal dimensions in prosody.

Not until recently has this two-sided face of prosody attracted the interest of the researchers of prosody. Models for duration and intonation were aimed at generating either the temporal or tonal structure. Over the past few years, however, some alleged connections of tonal movements with segmental lengthening have been reported, and a model which treats vowel

duration as a dependent variable of the change of Fo over this segment was presented by Lyberg (1981).

Yet, attractive as this idea might be at first glance and fruitful as it has been for prosodic research during the past few years, it has to be stated that Lyberg's model, in principle and in its absolute form, does not seem to hold against evidence of various kinds presented for Central Swedish (Bruce 1981), Central and Southern Swedish (Bannert 1982a), German (Bannert 1982b), and Danish (Thorsen 1980). It is of course true that nobody will doubt that there are certain dependencies between parts of the temporal and tonal structures. However, their mutual effects are rather small and their interrelationships should be characterized as mutual modification rather than dependence.

The second basic question concerns the input to the model, especially its phonological information. Which phonological features are to be contained in the input which is a linguistically abstract structure showing, among other things, semantic, syntactic, and pragmatic feature specifications? In this instance, too, different alternatives for a solution are possible and plausible, depending on the degree of redundancy that is to be allotted to the phonological structure.

For the present outline I assume that the input shows the following characteristics: under the syntactic category symbols, like N, ADJ, PRON, etc., discrete phonological elements (segments, phonemes) of the canonical lexical units are ordered linearly and specified as to their spectral features. Furthermore, the input contains the prosodic features STRESS and ACCENT of syllables, QUANTITY of vowels, and intonation type COMPLETED of the whole utterance. The input also contains information on various boundaries and the pragmatic features of CONTRAST of words and EMPHASIS of prosodic phrases and PHRASE INTONATION which signals the division of an utterance into minor prosodic units, the prosodic phrases, conditioned by several factors such as tempo, syntactic structure, etc. Thus the input, having the form of an abstract, linguistically defined structure, contains all the necessary phonological features, both spectral and prosodic,

and various boundaries. The prosodic features originate from lexical, morphological, syntactic, semantic, and pragmatic features and are inserted into the structure in their appropriate positions on different levels (segment, syllable, word, phrase, utterance, text).

As we can see, the input, as a linguistic and pragmatic structure, is abstract and therefore not specified in the temporal or tonal dimension. It is the task of the prosody model, by applying rules, to generate the concrete temporal and tonal structures in the time-fundamental frequency-field.

Accepting these assumptions, a possible conception of the prosody model is chosen which is shown in Figure 3. The model consists of three main components: the basic temporal and tonal components and the modification component. They may operate in the order mentioned. Thereby two aspects are stressed. Firstly, the temporal and tonal components operate separately, each of them generating a basic temporal and a basic tonal structure. Secondly, tonal features or gestures (tonal changes) like rises and falls or the points or levels HIGH and LOW, and the tonal range WIDE, only exist in reference to events in the time domain, i.e. tonal features are only meaningful if they are associated with temporal units like consonants, vowels, VC-boundaries or syllables.

The basic features of the prosody model outlined here are not unique for German but may also be valid for other languages. Prosodic features that are typical for German, however, are to be found in the phonology of German and in the rules proper which are contained in the temporal and tonal components (the re-writing rules and the intonation algorithm) and in the modification component.

It is the task of the modification component to adjust the basic temporal and tonal structures to their mutual characteristics and demands, and to adjust for assimilations and tempo constraints. Such adjustments are: temporal effects on the tonal contour, for instance the number of unaccentuated syllables between two accentuated ones; conversely, tonal effects on durations, for instance a rising tone may increase segment duration to a certain degree; the so-called micro-

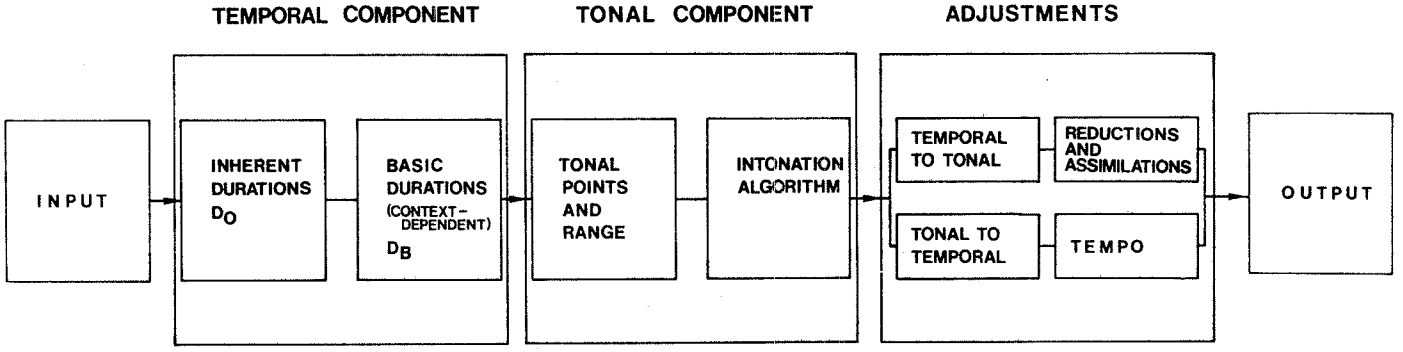


Figure 3. The basic structure of the model for German prosody.

prosody, the effects of the spectral features of the segments on the Fo-contour; speech tempo, reductions and assimilations causing rather drastic alterations in the prosodic structure (and, of course, in the spectral structure, too). It is, however, not self-evident to include tempo and assimilation in the modification component. They could operate separately as well. And, last but not least, a device is needed in order to introduce certain optional (individual) features into the prosodic structure.

As a concluding remark concerning general features of prosody models, I would like to point out another possible conception of a prosody model. Being totally different from the present outline, it does not generate the prosodic structure of an utterance in a linear, step-by-step fashion. Instead, this model derives the target structure in an open procedure, simultaneously on several levels, each level, at any moment, having access to all information on all the other levels. This hint at a dynamic sketch of a prosody model, however, will not be pursued here, although this concept obviously seems to be very attractive.

4. GENERATION OF THE BASIC TEMPORAL STRUCTURE

As I have followed a well-known approach[5], the generation of the basic temporal structure will be sketched here only very briefly. The segment is the basic unit of the temporal structure and its duration is calculated with respect to the different contexts and the domain it appears in.

In the temporal component of the model, the linguistically specified input is used to generate the basic temporal structure in two steps. Firstly, the discrete spectral units are assigned the inherent duration Do which is listed in the catalogue of basic durations. Which inherent durations are contained therein, has yet to be established (but cf. Rietveld 1975). In any case I assume that two temporal vowel categories have to be discerned as a consequence of the prosodic feature of QUANTITY, specified as ±LONG in the lexical units.

In the second step, the inherent duration Do may be modified as a consequence of certain context features. Such contexts

are the syllable features STRESS and ACCENT, phrase boundaries, consonant clusters, CONTRAST and EMPHASIS of words and phrases, position within stress groups and phrases. Each context may have the effect of increasing or decreasing the inherent durations, and the size of this change can be expressed by a factor. If several different context conditions are applicable to a given segment or syllable, the final size of the durational change is the sum of all these factors. Thus the basic duration D_B of a given segment is

$$D_B = (a + b + c + d + \ldots + n)\ Do$$

As was shown above (section 2.1), the intonation type has no systematic effect on sentence and stress group durations. Therefore, the context component contains no factor that would apply to a given intonation type.

The concrete basic temporal structure is then added to the original abstract structure of the input. Thus the structure which leaves the temporal component serves as the input to the tonal component. It contains the linguistic features of the input and the phonetic temporal structure in the time dimension.

5. GENERATION OF THE BASIC TONAL STRUCTURE

The generation of the basic tonal structure will be dealt with in more detail. This, in the first place, is due to some novel features in the intonation algorithm. The basic tonal component also operates in two steps[6]. Firstly, the tonal prosodic features, such as accent, intonation type, emphasis, etc., are transposed into a notation of tonal points or levels and of range. Secondly, operating on this information, the intonation algorithm then generates the tonal structure in the time-frequency-field in five steps, thus expanding the amount of concrete information present in the total structure of the given utterance.

5.1 The tonal re-writing rules

Apart from the rhythmic (temporal) prosodic features STRESS and QUANTITY of syllables and words, and different boundaries, at least the following tonal features have to be assumed for

German. From a phonetic point of view, they are described as tonal points or levels HIGH and LOW and as tonal range WIDE:

A. Tonal points or levels

±AKZENT: Accent, manifested in the fundamental frequency dimension, a tonal rise from an Fo-minimum to an Fo-maximum throughout the accentuated vowel. The final accent of a statement, however, appears as a fall, thus reflecting the superiority of the feature of intonation type (sentence intonation), namely a low tonal end, over the syllable or word feature of accent.

±ABGESCHLOSSEN: Completed, feature of the intonation type (sentence intonation), dominating the syllable or word feature of accent (and thus of a smaller domain, lower in the hierarchy), manifested in the fundamental frequency dimension, mainly local manifestation at the end of an utterance[7]: low in completed utterances (often statements), high in non-completed utterances (utterances where questioning is signalled by tonal means).

B. Tonal range

±KONTRAST: Contrast, concerns one word, pointing out one lexical element, manifested tonally by an enlarged range of the Fo-change in the accentuated syllable, accomplished by a heightened Fo-maximum.

±TEILNAHME: Involvement (emphasis), concerns the larger units of phrase and sentence, manifested tonally by enlarged ranges of the Fo-changes of all the accents, accomplished by heightened Fo-maxima.

In the first step of the tonal component, a re-writing algorithm transforms the specification of the tonal prosodic features into a notation consisting of the points HIGH or LOW and the range WIDE. The following rules are to be applied

(the arrow means "re-write as"):

A'. Tonal points or levels

+AKZENT → H(IGH) in the pre-accentuated consonant, applies to the final accent in completed phrases (statements) only.

→ L(OW) in the pre-accentuated consonant, otherwise.

+ABGESCHLOSSEN → L(OW) in the final vowel (or syllable)

-ABGESCHLOSSEN → H(IGH) in the final vowel (or syllable)

B'. Tonal range

+KONTRAST → W(IDE) in this LOW or HIGH

+TEILNAHME → W(IDE) in all non-sentence final LOWs and HIGHs.

By applying the tonal re-writing rules, a tonal specification is produced which also reflects the characteristic contrasts between the various prosodic conditions, for example intonation type and involvement. The operation of the tonal re-writing rules is illustrated by the following example taken from the material in the acoustic investigation. The sentence "Der Müller in Lingen will die Männer immer Lümmel nennen" containing four accents is shown as a statement and as an information question (inverted word order):

INPUT

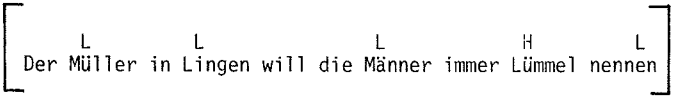
statement

[+ACC Der Müller +ACC in Lingen will die +ACC Männer immer +ACC Lümmel nennen] +COMPL

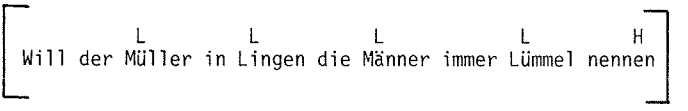
information question

[+ACC Will der Müller +ACC in Lingen die +ACC Männer immer +ACC Lümmel nennen] -COMPL

statement



information question



The tonal patterns, being the result of the tonal re-writing rules, are now altered into concrete Fo-contours by the intonation algorithm.

5.2 The intonation algorithm

The tonal patterns, specified in terms of L, H, and W and tied to spectral events in the time dimension, are transformed into Fo-values in the fundamental frequency dimension by the intonation algorithm, the second part of the tonal component of the model. This is done in five steps where the basic temporal structure is the necessary prerequisite for the operation of the intonation algorithm. Thus the abstract linguistic and the temporal phonetic structure (information) of the utterance is extended by the addition of the concrete tonal structure.

The intonation algorithm is designed in such a way as to generate the Fo-contour of an utterance from below, starting from the tonal floor of the contour, i.e. its Fo-minima, especially the absolute Fo-minimum.

The application of the intonation algorithm is shown in Figure 4. The same sentence as above "Der Müller in Lingen will die Männer immer Lümmel nennen" containing four accents of equal weight, to be spoken as a one-phrase statement and

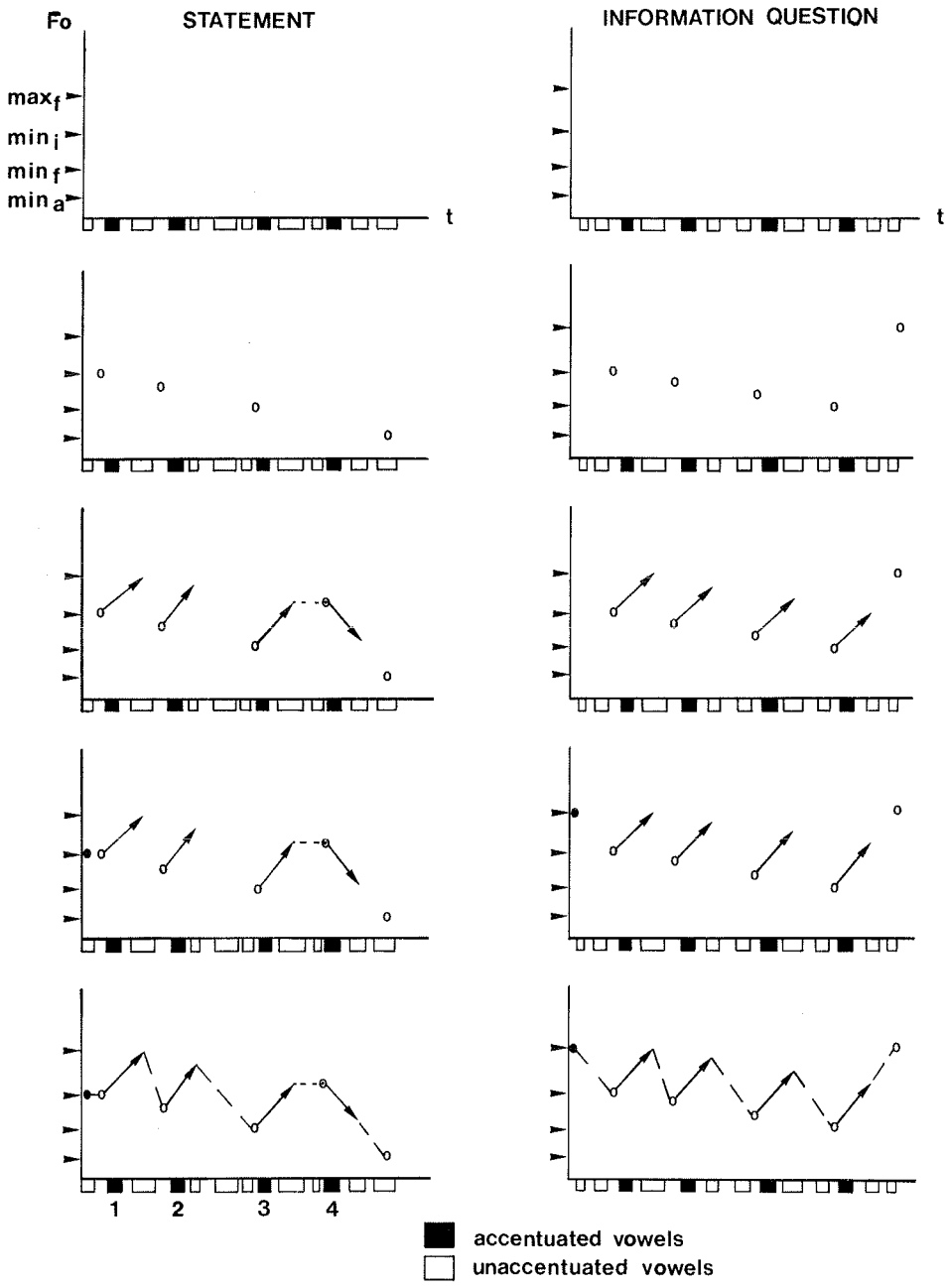


Figure 4. The intonation algorithm illustrating the generation of the tonal structure for a sentence with 4 accents as a statement and an information question.

information question, serves as an example. Where the accent pattern is concerned, this sentence corresponds to other sentences, for instance, "Das Gemälde von Kandinsky ist gestern versteigert worden."

STEP 1

In the first step, certain tonal targets (points or levels) have to be determined in order to fix the frame of the Fo-contour of the utterance. These target values are speaker-specific and have to be defined on an appropriate scale[8].

Three values correspond to accent minima (L), the fourth corresponds to the H of the question intonation:

1. $Fo\text{-}Min_a$: The absolute Fo-minimum, the lowest point at the end of a statement. It seems to represent the anchor of the whole tonal structure, the tonal movements (the tonal changes) and the tonal relationships within a given utterance.

2. $Fo\text{-}Min_f$: The final Fo-minimum, the lowest point of the final accent of questions.

3. $Fo\text{-}Min_i$: The initial Fo-minimum, the lowest point of the first accent. This Fo-value is highest for the statement and lowest for the echo question.

4. $Fo\text{-}Max_f$: The final Fo-maximum, the highest point at the end of questions. It may be equal to or higher than the first Fo-maximum.

The tonal points 3 and 1 or 2 respectively define the range of the Fo-minima, the bottom line, the tonal floor, throughout the prosodic phrase. All these points need not be reached in every utterance observed (cf. step 3). They function first of all as auxiliary points. As one global expression of the intonation type and involvement, point 3, the initial Fo-minimum of the first accent varies. Compared to statements, this Fo-minimum is lower in information questions and lowest in echo questions. As a result of this, the range of the bottom line, i.e. the declination, also varies.

STEP 2

In the second step, the fundamental points of the Fo-contour

are inserted. These are the Fo-minima of the accents and the end point of the utterance. The Fo-minima are located in the middle of the consonant preceding the accentuated vowel. First the following LOWs are specified from left to right in the utterance: the first LOW is given the value of $Fo\text{-}Min_i$ (point 3), the last LOW is given the value of $Fo\text{-}Min_a$ in statements (point 1) or the value of $Fo\text{-}Min_f$ in questions (point 2).

Second, assigning the intermediate Fo-minima, two cases are to be distinguished depending on the number of accents in the utterance:

1. Up to 5 accents: Divide the tonal interval between $Fo\text{-}Min_i$ and $Fo\text{-}Min_a$ or $Fo\text{-}Min_f$ into equal steps and insert these values into the intermediate pre-accentuated consonants.

2. More than 5 accents: Divide the tonal interval of the bottom line into equal steps. Insert these values, except the last but one, into the intermediate pre-accentuated consonants. Insert a higher value for the last but one accent.

In questions, the value of $Fo\text{-}Max_f$ is inserted for the final HIGH (point 4).

STEP 3

Upon this tonal floor, the tonal movements (rises) associated with the accents are inserted. Thus each Fo-minimum, except the $Fo\text{-}Min_a$ at the end of statements, represents the starting point of the tonal accent movement. Starting from a minimum, the movement can only go upwards, i.e. it must be a tonal rise, as the tonal floor is fixed.

These constant tonal movements are to be seen as tonal differences or tonal steps between syllables. Starting from an Fo-minimum in the consonant preceding the accentuated vowel, the tonal movement ends in the following, unstressed vowel (or in the CV-boundary). This is also true of the final accent of statements, although manifested as a tonal fall. The exact position of the end point of the tonal movement may vary, due to different temporal conditions such as the number of un-

stressed syllables following the accent or quantity (cf. Bannert 1982b).

There is, however, one accent movement that travels in the opposite direction, namely the final accent fall of statements. This final tonal fall has to be seen as a natural consequence of two factors: the high end point of the preceding accent rise and the low end point of the tonal contour in statements. Due to this tonal conditioning, the final accent fall can be derived easily using the tonal information available at this stage. As is apparent from the Fo-contours in Figure 2, the starting point of the final accent fall shows approximately the same Fo-value as the end point of the preceding accent rise[9]. Therefore this end point value is inserted into the final pre-accentuated consonant and the tonal movement is directed downwards. It terminates in the following unstressed syllable and shows the same speed of Fo-change as the rising movements.

The Fo-change in the accentuated syllable will be expressed by the factor $\underline{v}$ of a base unit, the absolute Fo-minimum (Fo-Min_a) which seems to be the lowest point or level of the normal speech register of a speaker. The factor $\underline{v}$, however, is not to be taken as an exact and fixed value; it is rather to be thought of as a certain Fo-range or interval from which values may be chosen for a given utterance.

If the specification WIDE appears with an accent, the size of the tonal movement $\underline{v}$ is increased by the factor $\underline{w}$ which also is speaker-specific and greatly variable. This factor, too, constitutes an Fo-range rather than a fixed value. Thus the total tonal movement V_G in an accent specified for LOW and WIDE is

$$V_G = (v + w)\ Fo_{Min_a}$$

The size of the tonal movements associated with the accents may vary throughout an utterance. This variation, it seems, expresses the concurrence of various features. In order to be heard as an accent, the tonal movement has to exceed a certain minimal value, a tonal threshold, defined from a perceptual point of view and dependent on certain contextual tonal cir-

cumstances. Larger tonal excursions of the accent movements are the result of such concomitant tonal features as contrast, emphasis, and phrase intonation. Therefore, the size of the tonal movements may be considered the sum of all contributions of tonal features and, as is the case with the basic durations (cf. section 4), may be defined as

$$V_G = (v + w + k + j + \ldots + n)\ Fo_{Min_a}$$

This means that the total tonal movement V_G associated with an accent is calculated with reference to the absolute Fo-minimum of the speaker. The proportion of the reference value is the sum of contributions of all the tonal features appearing together with a given accent: the minimal movement of the accent v, the range w of the feature WIDE reflecting involvement, etc. However, given equally strong accents in an utterance, I assume the tonal movements to be of equal size on an auditory scale.

It seems obvious that every change in the size of this equal and constant tonal movement signals some prosodic information, namely a deviation from the normal, i.e. expected situation of equal weight from accent to accent. A decrease of the tonal movement means a weakening of the accent decreasing the prominence of the semantic element; an increase, however, expresses a strengthening of the accent thus creating a greater prominence for this semantic element. Data concerning the tonal differences in the accents and the position of the Fo-minima suggest a strong interrelationship between these tonal features and the semantic structure of (parts of) the utterance[10]. Considering the utterances which contain the local adverbial in Lingen, it is found for all three speakers that the accent movement in Lingen is considerably larger than that of the accent in Müller. It is quite obvious that the latter accent appears somewhat weakened. This inequality between accents is shown in Figure 5 for each intonation type and for speaker B and is representative for all speakers. The tonal structures of two sentences are shown which contain seven accents but which differ in the position of the phrase boundaries: "Der lullende Müller# will die längeren Männer

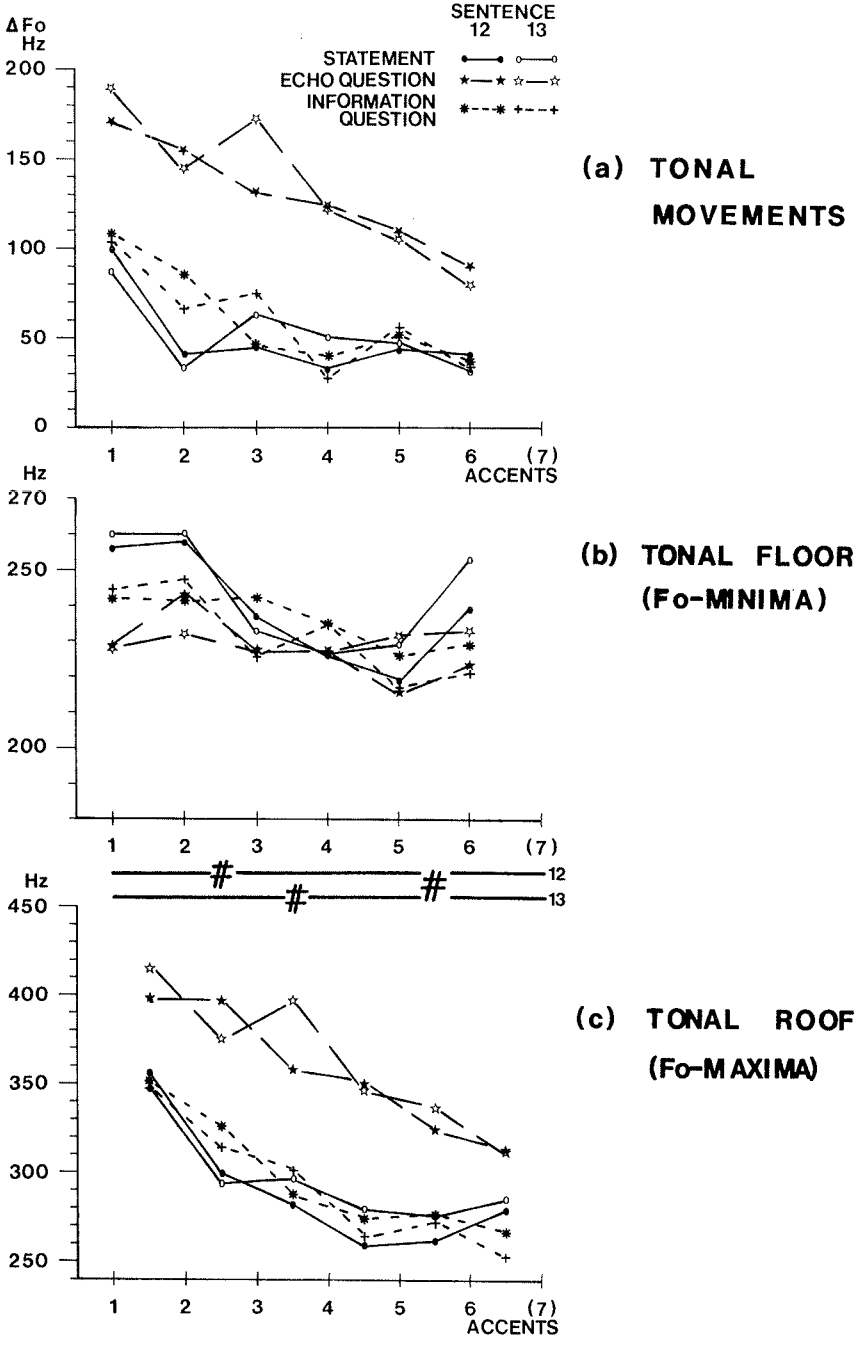


Figure 5. Variation of tonal characteristics due to syntactic and semantic factors. Speaker B. See text.

in der Menge immer lungernde Lümmel nennen" and "Der lullende Müller in Lingen # will die längeren Männer immer lungernde Lümmel nennen". The subject of the first sentence consists of a phrase containing two accents, that of the second sentence consists of a phrase containing three accents (cf. Table 1). In part (a) of Figure 5, the tonal differences (ΔFo) in each accent are shown, in part (b) the bottom line, the tonal floor (Fo-minima), and in part (c) the peak line, the tonal roof (Fo-maxima) are to be seen. It is evident that the tonal differences associated with the second and third accents (Müller and Lingen/längeren respectively) have implications for all three tonal characteristics, namely the Fo-movements, the bottom line, and the peak line. This entails a deviation from the gradually falling or uniform basic tonal structure.

STEP 4

To complete the tonal skeleton, the starting point of the whole Fo-contour is inserted. For statements and echo questions with an unstressed syllable preceding the first stressed one, a value near the initial Fo-minimum is chosen as a first, rough approximation. For the information question showing inverted word order, the starting point is chosen in relation to the initial Fo-maximum in such a way that the starting point equals the first Fo-maximum.

STEP 5

Finally, these fragments or basic elements of the tonal structure have to be connected together in order to arrive at the complete tonal contour of the given utterance. This can be done by straight lines as in Figure 4 which are smoothed out later on or by a cosine function.

The output from the intonation algorithm, the generated basic tonal contours, together with the basic temporal structure, will then be adjusted in various ways in the modification component. However, these adjustments are rather small where size and degree are concerned. The basic features of the Fo-contours, in principle, will not be altered. This means that, in spite of all modifications, the basic character and features of the prosodic structure are maintained.

6. THE MODEL AND INTONATION THEORY

The prosody model presented in this outline, after due amendments and completion, makes the claim of generating the temporal and tonal manifestations of the pure linguistic-prosodic features of a given German sentence, excluding the expression of emotions and other non-linguistic features. Operating on the abstract linguistic information of a sentence, the model components, step-by-step, will produce the concrete phonetic structure in the time and fundamental frequency dimensions.

The linguistic input of this prosody model contains all the prosodic features necessary to generate the prosodic structure of the given utterance. This does not mean, however, that there exists theoretical agreement about which prosodic features are to be assumed for German and how they are distributed in a sentence. Together with research into this issue from a purely general linguistic point of view (cf. Klein and von Stechow 1982, among others), the prosody model sketched here permits controlled experiments to be conducted in order to determine relevant prosodic features in a feed-back fashion. Even at this early stage, however, it seems appropriate to comment on some notions of intonation theory, especially for German as it is described by von Essen (1956).

There are no such independent features/elements like focus, sentence accent, "Schwerpunkt" (centre of gravity) or "progredient" intonation (cf. von Essen 1956) in this model. Rather they are dependent features since they are derived from the model elements[11]. It seems plausible, however, that there exists a difference between focus and contrast in German.

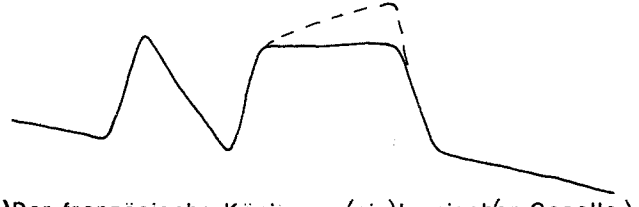
Focus or sentence accent or "Schwerpunkt" is an automatic result of accent and position. It always appears on the final accent of an utterance, be it a completed intonation (statement) or an uncompleted intonation (question). What is crucial, however, is that the following part of the contour remains undisturbed. No tonal movement (protrusion), except the gradually falling or rising declination, may occur after the focus movements. This final part of the utterance without any accent-like tonal movement is called "Nachlauf". From a seman-

tic point of view, it is very often a mere repetition of semantic elements and could readily be deleted. In other words, there are no accents following the sentence accent. Sentence accent is the final accent of a phrase, and together with the rest of the contour, also expressing intonation type, it is heard as "heavier, stronger" than the preceding, non-final accents.

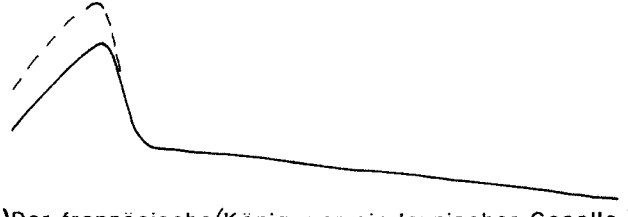
This effect is easily shown by manipulation of utterances by means of speech synthesis. For instance, starting from an utterance containing four accents of equal weight, e.g. "Der französische König war ein launischer Geselle" (The French king was a capricious fellow) and moving the final fall for the statement or the rise for the question in the accent positions to the left, while keeping the Fo-range of this final change constant, the focus (Schwerpunkt) is also moved to the left. The result of this manipulation is a shifting of weight or attention to different parts (semantic elements) of the utterances. Another effect is a shift of context for these utterances. This shifting of focus is illustrated in Figure 6a for statements and in 6b for questions.

In (1) of Figure 6a,the four accents of the statement are of equal weight. This statement may appear as the first sentence of a text, presupposing the general knowledge that there are kings of different countries. Although the accents are equal, each of them reflecting new information, the final accent is heard as the focus. This might be due to two circumstances. First, the final tonal movement (fall or rise) in a tone contour fulfills a particular role in signalling the intonation type and thus the global structuring of discourse. Second, on purely acoustic grounds, the tonal movement (being located in the lowest part of the tonal range of a statement) is perceived as large. The final rise constitutes a relatively large Fo-change from the lowest part of the Fo-range upwards. Both Fo-changes thus give rise to the auditory impression of a strong tonal element. The final tonal movement may be considered the combined result of two factors: the accent movement proper and the movement towards the very end point of the utterance.

(1) Der französische König war ein launischer Geselle.



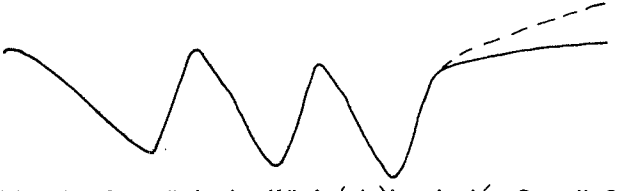
(2) Der französische König war (ein) launischer Geselle.)



(3) Der französische (König war ein launischer Geselle.)

Figure 6a. Shifting of focus (Schwerpunkt) and contrast from right to left in a statement.
Underlined focus word: ——
Underlined contrast word: – – –

(1') War der französische König ein launischer Geselle?



(2') War der französische König (ein) launischer Geselle?)

(3') (War) der französische (König ein launischer Geselle?)

Figure 6b. Shifting of focus (Schwerpunkt) and contrast from right to left in a question.

In (2) the focus is on launisch, in (3), however, on französisch. Shortened (elliptic) but equivalent versions of these two statements would be for (2) "Der französische König war launisch" and for (3) "Der französische". The repetitive parts of the original complete sentences are put into parantheses in Figure 6. The shortened versions could be uttered in the following contexts: "Was für ein Geselle war der französische König?" for (2), where the quality of the French king is in the semantic focus, and "Welcher König war ein launischer Geselle?" for (3), where the nationality of the king who was a capricious fellow is in the centre of interest. Correspondingly, the position of sentence accent is shifted to the left in questions (1'), namely to launisch in (2') and to französisch in (3').

In contrast to the sentence accent, being the utterance-final instance of several equal accents, a semantic element may appear under contrast, which means that this element is particularly pointed out by the speaker either as a paradigmatic device expressing a meaning not expected or unusual in this position or as a correction of a meaning in the answer to a statement previously uttered by the listener. The accent of this emphasized semantic element is stronger than the preceding ones, if there are any. Contrast on a word (a lexical element) is signalled tonally by a larger Fo-change associated with the word accent which is achieved by a higher Fo-peak. This is illustrated by the dashed line in Figure 6a, b (cf. also Bannert 1982b, Figure 1). A contrasting accent always shows up as the last one in a phrase; there can be no pitch protrusion in the Fo-contour following the contrastive accent. In order to bring out the contrast more clearly, the preceding accents, if any, may be weakened.

If the whole utterance is to be stressed, this will be done by means of emphasis, i.e. the whole utterance (its accents actually) will be spoken with high involvement from the part of the speaker. Contrast and emphasis are expressed by the same tonal means, namely increased tonal range, i.e. larger tonal movements in the accents, achieved by raising the Fo-maxima (cf. Figure 2). The only difference between these two features, besides their function, is their domain. Contrast,

emphasis, and sentence accent are illustrated in Table 2.

In my intonation model, the so-called "progredient" intonation (cf. von Essen 1956) does not have the same status as the intonation types ±COMPLETED which is the tonal feature of a whole utterance containing one or more prosodic phrases. "Progredient" intonation exists in certain contexts only: it never appears finally in an utterance and it is optional. This intonation seems rather to be a device for structuring an utterance by dividing it into smaller units (prosodic phrases) and thus signalling continuation within an utterance. "Progredient" intonation also seems to be associated with a clause boundary which also may be expressed by temporal and phonatory means. Therefore, in my model, "progredient" intonation corresponds to the feature of phrase intonation which is connected with the phrasing of an utterance. However, the tonal expression of phrase intonation is not yet quite clear, although there exist some data in Delattre et al. (1965).

Different strategies for the tonal expression of phrase intonation can be conceived of, namely a local one and a global one, and of course a combination of both. Locally, there may be a tonal jump from the high Fo-contour to the left of the phrase boundary down to the low Fo-contour to the right of the phrase boundary. In this case the abrupt tonal jump from one phrase to another contrasts with the gradual fall between successive accents within a phrase (cf. Bannert 1983a, Figure 1). Globally, there may be a larger tonal movement in the phrase-final accent, which, compared to all the accents in the whole utterance, expresses the division of the utterance into phrases in a complex way. It seems plausible that phrase intonation marks the subordination of syntactic units within a grammatically complex sentence. These two basic strategies are illustrated in Figure 7[12].

The model also accounts for the devices of accent weakening (backgrounding) and deaccentuation (cf. Bannert 1983a).

It is quite obvious that a considerable amount of work remains to be done in order to develop a complete model for German prosody, including Standard German and the dialects. Nevertheless my outline of a model for German prosody does

Table 2. Illustrations of focus (sentence accent, "Schwerpunkt"), contrast, and emphasis in statements in German. ' = lexical accent (accentuated word), * = accent of involvement (contrast on one word, emphasis on the whole phrase). ( ) = repetitive parts of utterances, usually dropped; / = possible substitution.

Situation:	Zwei Nachbarsfamilien A und B treffen sich an einem Samstag vormittag auf dem Parkplatz eines Einkaufszentrums. Es ist Frühling; die Gartenarbeit wartet. In den Nachbardörfern Schwanfeld und Moosbach ist jeweils ein Gärtner.	
SCHWER-PUNKT:	Dialog 1: A: Na, was habt Ihr denn heute vor?/Wo geht's denn hin? B: (Wir wollen) zum 'Gärtner in 'Schwanfeld (fahren).	*Beide Akzente gleichwertig. Schwerpunkt automatisch auf dem letzten Akzent.*
	Dialog 2: A: Margit hat vorgestern gesagt, Ihr wollt heute nach Schwanfeld fahren. B: Ja, (wir wollen) zum 'Gärtner (in Schwanfeld) fahren.	*Schwerpunkt auf dem letzten (einzigen) Akzent.*
KONTRAST:	Dialog 3: A: Fahrt Ihr zum Gärtner in Moosbach? B: Nein, (zum 'Gärtner/zu 'dem) in *Schwanfeld!! Ihr wisst doch, der in Moosbach ist viel zu teuer!!	*Korrektur einer Annahme oder verwunderter Ausdruck, betrifft ein Wort (einer Phrase).*
	Dialog 4: A: Fahrt Ihr zum Schmied in Schwanfeld? B: Nein, zum *Gärtner (in Schwanfeld).	*Schwanfeld deakzentuiert, da es auf das kontrastierende Wort folgt. Identisch mit Dialog 6.*
EMPHASE:	Dialog 5: A: Fahrt Ihr zum Schmied in Moosbach? B: Nein, zum *Gärtner in *Schwanfeld!! Habt Ihr schon wieder vergessen?! Wir haben Euch doch erst gestern gesagt, dass wir heute zum Gärtner in Schwandorf fahren!!	*Ausdruck der Teilnahme (Irritation, Verärgerung), betrifft die gesamte Phrase (mit mehreren Akzenten).*
	Dialog 6: A: Fahrt Ihr zum Schmied? B: Nein, zum *Gärtner!! Habt Ihr denn schon wieder vergessen?! Wir haben Euch doch erst gestern erzählt, dass wir heute zum Gärtner fahren!!	*Ausdruck der Teilnahme (Irritation, Verärgerung), betrifft die Phrase, hier mit nur einem Akzent. Identisch mit Dialog 4.*

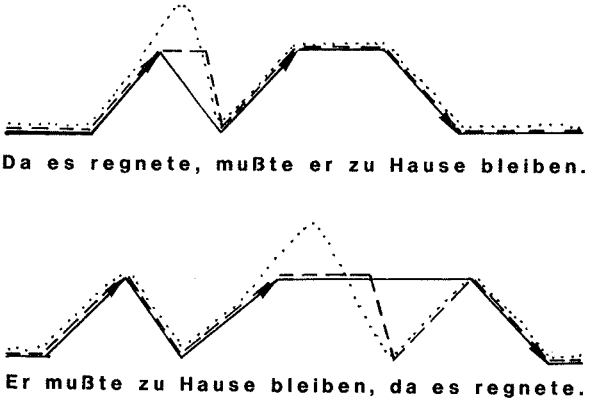


Figure 7. Hypothesized Fo-contours for phrase intonation.

contain some basic aspects relevant to this field of research. Further research work will bring us closer to an understanding of the workings of prosody and thus better qualify us for making intelligible and natural speech synthesis and enable us to more effectively teach German as a foreign language. Future prosodic research will include language-specific German features of Standard German and of the dialects along with features that German shares with other related and non-related languages. As a particular challenge, the close relationship between prosody and syntax and semantics (pragmatics) hopefully will promote the cooperation between phoneticians, general linguists, and psychologists. With these perspectives in mind, we are justified in looking forward to some very exciting and productive research work on German prosody.

FOOTNOTES

1 The book by Isačenko and Schädlich (1970), a translation of their paper (1966), is a pure description of simplified tonal conditions. In spite of its title, it does not contain any rules or algorithms which generate tonal structures.

2 A shortened version of this paper was presented at the Interdisciplinary Symposium "Prosody - Normal and Abnormal" held at the University of Zürich, April 6-8, 1983 and arranged by the Association Européenne Psycholinguistique. It is not the aim of this paper to compare the model proposed here to other models to be found in the literature. For a preparatory work, focusing on the intonation part, see Bannert (1983a).

3 The phonetic aspects, particularly the acoustic prerequisites for the conception of the intonation algorithm and the details of the investigation, are dealt with in Bannert (1983b).

4 Choosing the test sentences, the phonetic demands on a material which is heavily controlled for linguistic and acoustic variables dominated over semantic aspects. Although the test sentences do not constitute the most frequent German utterances, they do form possible linguistic structures. The assumption underlying the strong phonetic choice is represented by the claim that the basic linguistic elements of speech have to be known before it can be meaningful to study the prosody of spontaneous speech with all its numerous variables, the effect of which is not yet known. The longest utterance, containing eight accents, reads in English "The miller in Lingen doing pee-pee will always call the taller men in the crowd loitering louts".

5 The solution adopted here corresponds to the model for segment duration presented by Klatt (1979) and Lindblom et al. (1981), to mention only a few. Exactly the opposite

procedure, working in a top-down fashion, i.e. starting from the largest unit, the sentence, and going down via stress group and syllable to the segments, is to be found in Kohler (1982) and Kohler et al. (1982). Neither will the consequences of the so-called gesture theory suggested by Öhman et al. (1979) be considered here (but cf. Bannert 1982a).

6 The Lund model of intonation (Bruce 1977, Gårding and Bruce 1981) served as a special source of inspiration in outlining the basic tonal component. However, several significant differences, not due to language differences between Swedish and German, were introduced into the model for German prosody. I felt a need to replace the visually-oriented and geometrical conception of lines (base and top line, focal lines) by a dynamic, listener-oriented view of tonal changes and tonal relations (cf. also Bruce 1982).

7 The prosodic primes "utterance" and "phrase" are assumed as given units. Nothing concerning the difficulty of defining them will be discussed here.

8 Mean values for five tonal points in the three intonation types of three speakers are to be found in Bannert (1983b, Table 3).

9 This derivation in statements is only possible if the utterance contains at least two accents. In a statement containing only one accent, the tonal point HIGH of the final and, at the same time, first accent has to be inserted directly by the intonation algorithm.

It cannot be taken for granted that the hat pattern constitutes the only and obligatory tonal contour of the two final accents in statements. Optionally the rise of the last but one accent may be followed by a fall and rise preceding the final accent fall. Thus each accent is characterized by a tonal rise-fall movement, although the tonal gesture is timed differently (earlier) for the final accent. Evidence for this tonal solution is to be found

in the Fo-tracings of speaker E.

10 A more detailed account is to be found in Bannert (1983b).

11 This section is based on, among other things, the results of experiments using LPC-synthesis where various utterances of the speakers B and K were manipulated according to the model elements. Thus, by just changing the Fo-contour locally at the end of an utterance, in accordance with the model, a statement is changed into an echo question and vice versa. If the final fall or rise in utterances containing several accents is moved towards the beginning of the utterance, the syllable having received the final tonal movement appears as the "Schwerpunkt" or sentence accent. However, systematic listening tests with naive listeners need to be run.

12 This presentation is based on the Fo-contours of the author. They were observed on a storage oscilloscope connected to a pitch meter. It remains to be investigated, too, how the clause boundary is expressed temporally, for instance by means of final lengthening.

REFERENCES

Bannert R. 1982a. An Fo-dependent model for segment duration? Reports from Uppsala University Department of Linguistics 8, 59-80

Bannert R. 1982b. Temporal and tonal control in German. Working Papers 22, 1-26. Department of Linguistics and Phonetics, Lund University

Bannert R. 1983a. Modellskizze für die deutsche Intonation. Zeitschrift für Literaturwissenschaft und Linguistik 49, 9-34

Bannert R. 1983b. Some phonetic characteristics of a model for German prosody. Working Papers 25, 1-34. Department of Linguistics and Phonetics, Lund University

Bolinger D. 1958. A theory of pitch accent. Word 14, 109-149

Bruce G. 1977. Swedish word accents in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII. Lund

Bruce G. 1981. Tonal and temporal interplay. Working Papers 21, 49-60. Department of Linguistics and Phonetics, Lund University

Bruce G. 1982. Developing the Swedish intonation model. Working Papers 22, 51-116. Department of Linguistics and Phonetics, Lund University

Delattre P., E. Poenack, and C. Olsen. 1965. Some characteristics of German intonation for the expression of continuation and finality. Phonetica 13, 134-161

Essen O. von 1956. Grundzüge der hochdeutschen Satzintonation. Ratingen

Gårding E. and G. Bruce. 1981. A presentation of the Lund model for Swedish intonation. Working Papers 21, 69-75. Department of Linguistics and Phonetics, Lund University

't Hart J. and A. Cohen. 1973. Intonation by rule: a perceptual quest. Journal of Phonetics 1, 309-327

Isačenko A.V. and H-J. Schädlich. 1966. Untersuchungen über die deutsche Satzintonation. Studia Grammatica VII, 7-67

Isačenko A. and H-J. Schädlich. 1970. A Model of Standard German Intonation. The Hague

Klatt D. 1979. Synthesis by rule of segmental durations in English sentences. Proceedings of the 9th International Congress of Phonetic Sciences. Vol. II, 290-297. Copenhagen

Klein W. & A. von Stechow. 1982. Intonation und Bedeutung von Fokus. Sonderforschungsbereich 99 Linguistik, University of Konstanz

Kohler K.J. 1982. Rhythmus im Deutschen. Arbeitsberichte 19, 89-105. Institute of Phonetics, University of Kiel

Kohler K.J., U. Krützmann, H. Reetz, and G. Timmermann. 1982. Sprachliche Determinanten der signalphonetischen Dauer. Arbeitsberichte 17, 1-48. Institute of Phonetics, University of Kiel

Lindblom B., B. Lyberg, and K. Holmgren. 1981. Durational patterns of Swedish phonology: Do they reflect short-term memory processes? Linguistics Club, Indiana University

Lyberg B. 1981. Temporal properties of spoken Swedish. Monographs from the Institute of Linguistics 6, Stockholm University

Öhman S., S. Zetterlund, L. Nordstrand, and O. Engstrand. 1979. Predicting segment durations in terms of a gesture theory of speech production. Proceedings of the 9th International Congress of Phonetic Sciences. Vol. II, 305-311. Copenhagen

Rietveld A.C.M. 1975. Untersuchung zur Vokaldauer im Deutschen. Phonetica 31, 248-258

Thorsen N. 1980. Neutral stress, emphatic stress, and sentence intonation in Advanced Standard Copenhagen Danish. Annual Report from the Institute of Phonetics, University of Copenhagen 14, 121-205

Thorsen N. 1982. Two issues in the prosody of Standard Danish: The lack of sentence accent and the representation of sentence intonation. Annual Report of the Institute of Phonetics, University of Copenhagen 16, 19-36