Working Papers 25 1983
Linguistics-Phonetics
Lund University, 1-34

SOME PHONETIC CHARACTERISTICS OF A MODEL FOR GERMAN PROSODY

Robert Bannert

ABSTRACT

In this paper some phonetic characteristics of a model for German prosody are described. The acoustic data form the most important basis for the outline of the model that is presented in Bannert (1983). The variables investigated are number of accents (sentence length), intonation type, position of phrase boundaries, and speaker. Aspects of the basic temporal and tonal components are illustrated. Variations of sentence and stress group durations do not show any clear and consistent pattern. The fundamental building blocks of the intonation algorithm, namely the basic tonal points, the tonal floor and the tonal movements in the accents, are treated in more detail. The effects of different factors on these tonal movements become evident.

1. BACKGROUND AND AIM

Compared to other languages and to my knowledge, no model existed until recently which was able to generate the prosody, i.e. the rhythm and the melody, of German utterances. Lately, however, an attempt has been made to fill this gap. An outline of an intonation model for German was presented in Bannert (1983a) and a model for German prosody was sketched in Bannert (1983b).[1]

The basic concept of the model for German prosody, which in its main components may also be valid for other languages, is shown in Figure 1. The model which is meant to generate the temporal and tonal structure of utterances consists of three main components, namely the basic temporal and tonal components and the modification component. The input to the

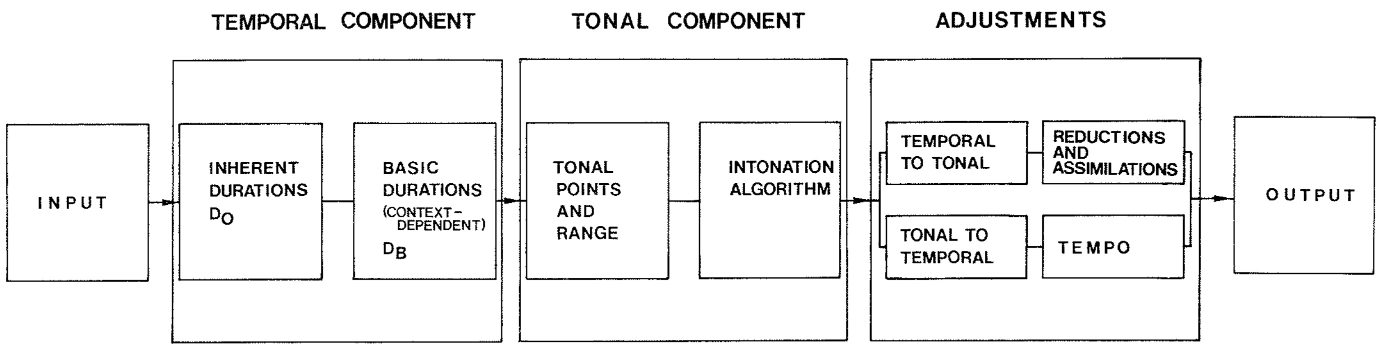

Figure 1. The three main components of a model for German prosody.

model is an abstract, linguistically completely specified structure. The output is the corresponding concrete temporal and tonal structure, i.e. the durations of segments and the Fo-contour.

The outline of the prosody model is based on, among other things, a relatively large acoustic material. However, in the papers mentioned above, the acoustic data could not be treated. Therefore, the aim of this paper is to account in some detail for the phonetic characteristics of the prosody model, especially the duration and intonation components.[2] This will be done by presenting the observed German data against the background of some general and important problems of a prosody model.

In order to provide the phonetic frame, the investigation is presented first. Secondly, the temporal structure on the sentence and stress-group level is illustrated. Finally, the tonal data are treated in more detail.

When evaluating the data an attempt is made to draw a distinction between linguistic features which are obligatory for all speakers of the language and individual features which are typical of a certain speaker and which therefore are optional. In the basic components of the model only such features should appear which are elements of the common linguistic code. Speaker-specific features may be incorporated into the model on the individual level. As the criterion to determine the limit between common linguistic and individual prosodic features in the material, the claim is made that a certain observation has to be found with all three speakers in order to qualify as an obligatory feature.

2. THE INVESTIGATION

The investigation was confined to the unit of the prosodic phrase, i.e. the utterances had to be spoken as one breath group and the accents should receive equal weight. Starting from this restriction, it seemed desirable to consider very different and almost extreme conditions in order to obtain basic acoustic data to build the model upon. Therefore, the following variables were chosen:

1. Length of utterance (prosodic phrase): number of accents 1 - 8, total number of syllables 3 - 29.
2. Intonation type: falling vs. rising melody at the end of the utterance, here called statement vs. echo and information question for the sake of simplification.
3. Speaker's involvement: normal, calm, detached (statement, information question) vs. involved, surprised, excited speech (echo question).
4. Syntactic structure: position of syntactic boundaries between accents (number of accents per syntactic phrase).
5. Number of speakers: three.

Table 1 shows the material which consisted of 14 utterances each of which was elicited as the intonation types statement, echo question, and information question.[3] Utterances 2-13 are pairs containing the same number of accents, but differ with regard to the position of the syntactic phrase boundaries. Thus, for instance, in sentence 4 (with 3 accents) the three accents make up one syntactic phrase. In sentence 5, however, the three accents are to be found in three syntactic phrases.

The sentences were constructed with regard to phonetic features which, of course, very often results in utterances that do not belong to every day language use. Nevertheless, these utterances are possible choices in the language (as to the conditions of the choice of the material, cf., for instance, Bruce 1977). Each accentuated vowel of the sentences is phonologically short and not open. Most of the consonants are sonorants, and thus hardly affect the Fo-curve.

The sentences were read by three female speakers from Northern Germany (speaker B, born 1956 in Lübeck, speaker K, born in 1955 in Uetersen, near Pinneberg, and speaker E, born in 1957 in Meldorf in Dithmarschen). The speakers, who were enrolled at the University of Kiel, were instructed to read the sentences without hesitation or break or pause and to give equal weight to each content word. The speakers chose their own individual speech tempo.

The material was recorded in the anechoic chamber of the Institute of Phonetics at the University of Kiel[4] on a Revox tape recorder, operating speed 9 ips. The three intonation

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															
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.
						#						#				

types statement, information question, and echo question were presented in appropriate contexts and treated in three blocks in the order mentioned. The informants knew the little story about an article in a newspaper which contained all the 14 test sentences.[5] They were also well acquainted with the chamber and the recording procedure.

The method of eliciting the three intonation types is illustrated by the following examples (cf. Bruce 1977):

1. Statement. As the answer to an appropriate question, the statement was written on a card, e.g.

> 31
>
> Welche Nachricht aus unserer Stadt brachte der Kurier heute?
>
> Der Müller will die Männer immer Lümmel nennen.

In a kind of dialogue, the speaker would first read the reference number, then the question, and finally the answer (the statement).

2. Information question. The question was written on cards as a response by the speaker to a statement which referred to the known article in the newspaper "Kurier", e.g.

> 31
>
> Die Geschichte in der Zeitung machte die Leute neugierig.
>
> Will der Müller die Männer immer Lümmel nennen?

3. Echo question. The story was read aloud by the test leader. The speaker pronounced each test sentence which was underlined in the text by repeating it in the same word order as the statement in the text, but, at the same time, expressing involvement, namely surprise or astonishment.

Each sentence was spoken in five series. The recorded material was checked auditorily as to the correct realisation of the intonation type and the accents. For the acoustic analysis the material was processed by a Frøkjaer-Jensen pitch meter in the laboratory of the Department of Linguistics and Phonetics at Lund University. Duplex-oscillogrammes and Fo-curves were analysed and measured by hand. Each value to be presented later represents the arithmetic mean of five measured values.

3. DURATIONS

In the second step of the basic temporal component (cf. Figure 1) the inherent segment durations are lengthened or shortened due to various contextual effects. In the following account of some temporal aspects it will be shown how the durations of sentences and stress groups behave in the intonation types statement, information question, and echo question. Clear patterns of temporal variation will indicate the need to incorporate the context factor "intonation type" into the second step of the basic temporal component of the model. Thus the following questions will be asked:

1. Do the three intonation types show their own characteristic durations? Are there any clear and consistent differences between the sentence durations of the intonation types?
2. How do the stress group durations vary as a consequence of intonation type? And if they do, is this difference located at the end of the sentences (the last stress group) where it could be a consequence of the opposed Fo-contours (fall vs. rise)?

An account of the temporal features on the levels of syllables or segments will not be given in this paper.

3.1 Sentences

Means and standard deviations (in ms) of the duration of each of the 14 sentences according to intonation type and speaker are given in Table 2. An examination must be made as to whether the sentence durations of the three intonation types are different from each other or not. As a simplified measure, the threshold value of the durational difference ΔD_S is chosen according to the following definition: The sentence duration of two intonation types is clearly different if the durational difference ΔD_S is greater than the sum of their standard deviations s_1 and s_2, thus

$$\Delta D_S > s_1 + s_2$$

If the duration of the statement is taken as the reference of this comparison, the two following conditions hold:

Table 2. Mean sentence durations and standard deviations (ms) for each intonation type (I,A,E) and for each speaker (B,K,E). First line $\bar{x}$, second line s. A clear difference between durations is indicated by an asterisk (see text).

sentence number	number of accents	speaker:	B			K			E		
		intonation type:	I	A	E	I	A	E	I	A	E
1	1		448	* 498	* 608	470	448	* 590	378	* 440	410
			19.2	23.9	39.6	10.0	19.2	49.0	11.0	24.5	25.5
2	2		828	* 890	* 998	798	822	* 940	780	838	842
			22.8	35.4	56.8	25.9	22.8	62.0	51.0	34.2	28.6
3	2		966	1008	* 1170	862	928	* 1050	888	* 1006	986
			39.7	22.8	27.4	37.0	31.1	67.8	40.9	36.5	27.0
4	3		1368	1376	* 1478	1228	* 1298	* 1487	1270	* 1394	1384
			42.1	27.0	40.9	37.7	27.7	89.6	32.4	36.5	64.3
5	3		1884	* 1972	1962	1654	1700	1813	1694	1764	1764
			27.0	19.2	55.0	28.8	30.8	102.1	32.1	68.0	45.6
6	4		2250	2282	* 2438	2008	2036	2210	2058	-	2174
			43.6	42.1	76.3	35.6	69.1	116.8	55.6		46.2
7	4		2294	2350	2404	2000	* 2138	2250	2036	* 2234	2192
			29.7	51.5	110.6	46.4	56.3	195.7	61.1	65.8	49.2
8	5		2700	2764	* 2900	2322	* 2604	2648	2514	* 2660	2574
			79.1	65.0	56.6	53.6	129.7	149.6	75.0	60.4	61.1
9	5		2798	2860	* 3014	2496	* 2608	* 2756	2606	2716	2798
			64.2	68.6	74.0	32.1	39.6	86.8	139.6	82.6	62.2
10	6		3116	3112	3172	2764	2856	3002	2850	-	3078
			71.6	74.0	108.3	58.6	51.3	132.2	31.6		61.4
11	6		3156	3132	* 3286	2826	2918	2980	2944	3044	3108
			59.4	13.0	129.9	65.0	106.6	137.8	89.9	62.3	62.6
12	7		3694	3636	* 3898	3318	* 3536	3642	3490	3594	3654
			88.2	43.4	66.5	50.7	99.6	149.7	70.7	85.3	62.7
13	7		3578	3566	* 3810	3340	3376	3510	3398	3476	3526
			51.7	38.5	145.3	56.1	65.0	187.7	53.6	143.3	101.4
14	8		4084	4010	* 4270	3780	3796	4028	3874	4008	4098
			76.4	58.3	59.2	100.0	32.1	128.9	81.1	69.1	55.4

$$|D_A - D_I| > s_A + s_I$$

$$|D_E - D_A| > s_E + s_A$$

where A, E, I means the intonation type statement, echo question, and information question respectively, D is the sentence duration, and s the standard deviation.

All the cases, where the information question has a shorter duration than the statement, and the echo question has a longer duration than the statement, are marked with an asterisk between the duration values of the two intonation types which are compared in Table 2. It may be seen very clearly that no regular pattern among the three speakers is to be found. In total, clear differences of sentence duration between intonation types show up in only 35% of the 84 comparisons possible. Furthermore, the share for each speaker is very different: the echo question of speaker B nearly always has a longer duration than the statement. However, both have equal durations for speaker E. For the remaining comparisons, clear differences are to be found only in one third or one fifth of all cases. In order to simplify further, the relationship of the sentence durations between the three intonation types for each speaker may be expressed like this:

speaker	intonation type
B	I = A < E
K E	I = A = E

An irregular pattern of temporal behaviour of sentence durations is also to be found when another measure is used, namely the mean percentage of durational differences between each sentence and the intonation types. The mean values (%) of the 14 sentences are as follows:

speaker	A - I	E - A
B	0.07	7.4
K	3.6	8.6
E	6.3	-0.3

The mean durational differences between the intonation types compared do not exceed 10% and vary considerably between speakers and intonation types. According to this measure, the relationship between the sentence duration of the intonation types is expressed as follows:

speaker	
B	I = A < E
K	I < A < E
E	I < A = E

Figure 2 shows the pattern of variation of sentence durations between the intonation types for each speaker with the duration of the statement as the reference. For the sentences containing 2 to 7 accents, two curves are shown (cf. Table 1). It is evident from Figure 2 that no overall pattern of relative sentence duration is to be found. As the only common feature, it can be observed that the durational differences between intonation types are largest among the sentences containing the smallest number of accents.

These results concerning the sentence durations of the intonation types are relevant for the basic temporal component of the model. As the basic components should contain linguistic and general features only, i.e. features that are shared by all speakers, the second step of the basic temporal component will not include a context factor "intonation type" obligatory for all speakers.

3.2 Stress groups

As was the case with sentence duration, the durations of the stress groups do not show any clear pattern of variation either. Due to word inversion in the information question, the stress group durations of this intonation type could not be calculated. Thus stress group durations of statements and echo questions only are considered.

Figure 3 shows the stress group durations of statement and echo question of sentence 5 with 3 accents, sentence 9 with 5 accents and, sentence 14 with 8 accents (cf. Table 1) for each speaker. It can be seen that the pattern of durational variation is similar for all speakers. Thus two things be-

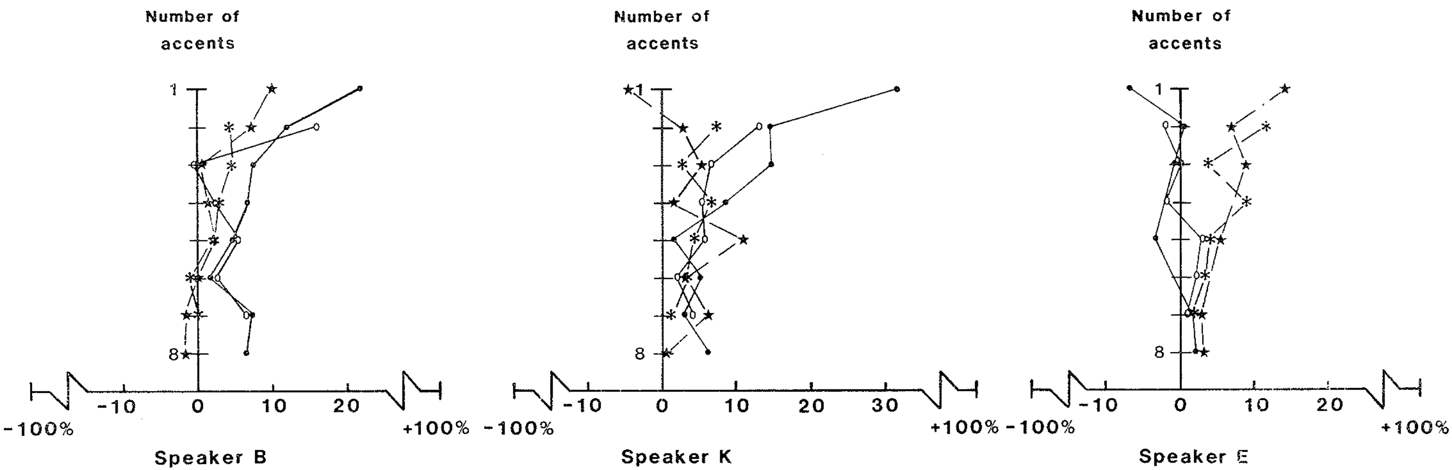

Figure 2. Relative mean differences of sentence duration between statement - information question (★,✻) and echo question - statement (●,o) for each speaker, the durations of statements being the reference.

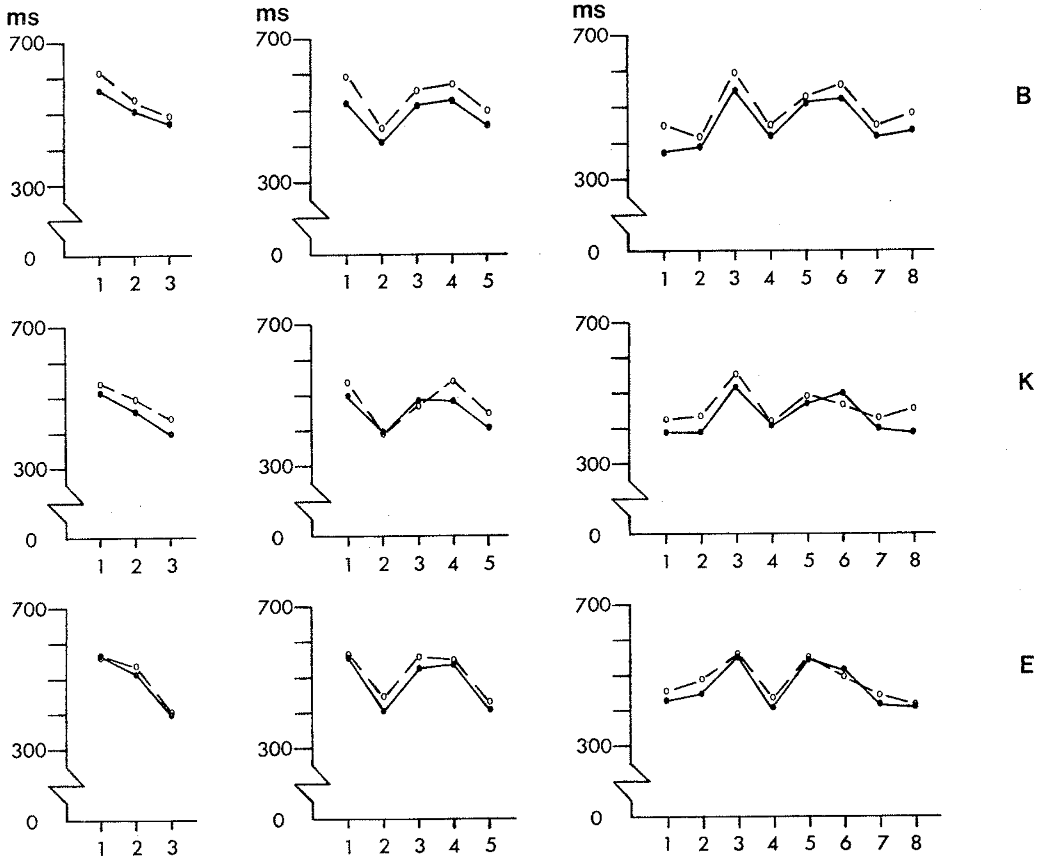

Figure 3. Mean durations of stress groups in sentences 5, 9, and 14 with 3, 5, and 8 accents respectively for each speaker (B,K,E). ●——● STATEMENT, ○— —○ ECHO QUESTION

come evident: First, stress group duration is determined by the duration of their elements, namely the syllables and segments. Evidence for this interpretation is to be found in the parallel variation of stress group durations for all speakers. Therefore it is taken to reflect a linguistic, i.e. general and obligatory feature of the temporal structure of utterances. Secondly, the difference in sentence duration reported above seems to be a global change of duration, i.e. it is spread over each stress group, and thus is not a consequence of the final tonal contours or the varying range of the tonal movements of the whole utterance (cf. Figure 5).

Figure 4 shows the relative differences of stress group duration between echo question and statement (= reference) of sentence 8 with 5 accents and sentence 14 with 8 accents. The total mean of each speaker is also included.

In parallel with the relative differences of sentence duration between the intonation types (cf. Figure 2), no regular pattern of durational variation can be discerned with the stress groups either. As a consequence, therefore, the basic temporal component of the prosody model will not contain any factor that may determine differences between various intonation types. The temporal component operates with one basic temporal structure, irrespective of intonation type. The durational differences, however, are not skipped altogether. They are seen as individual, and thus optional, features and are accounted for in the modification component by a device generating speaker-specific variation.

4. INTONATION

The basic tonal component of the prosody model consists of two parts, the tonal re-writing rules and the intonation algorithm (cf. Figure 1 and Bannert 1983a). Before data concerning the algorithm will be presented, some basic information on the Fo-contours will be given.

4.1 Fo-contours

Figure 5 shows the normalized Fo-contours of the utterances 1, 2, 5, and 14 containing 1, 2, 3, and 8 accents defined by the initial Fo, the Fo-minima, the Fo-maxima, and the final

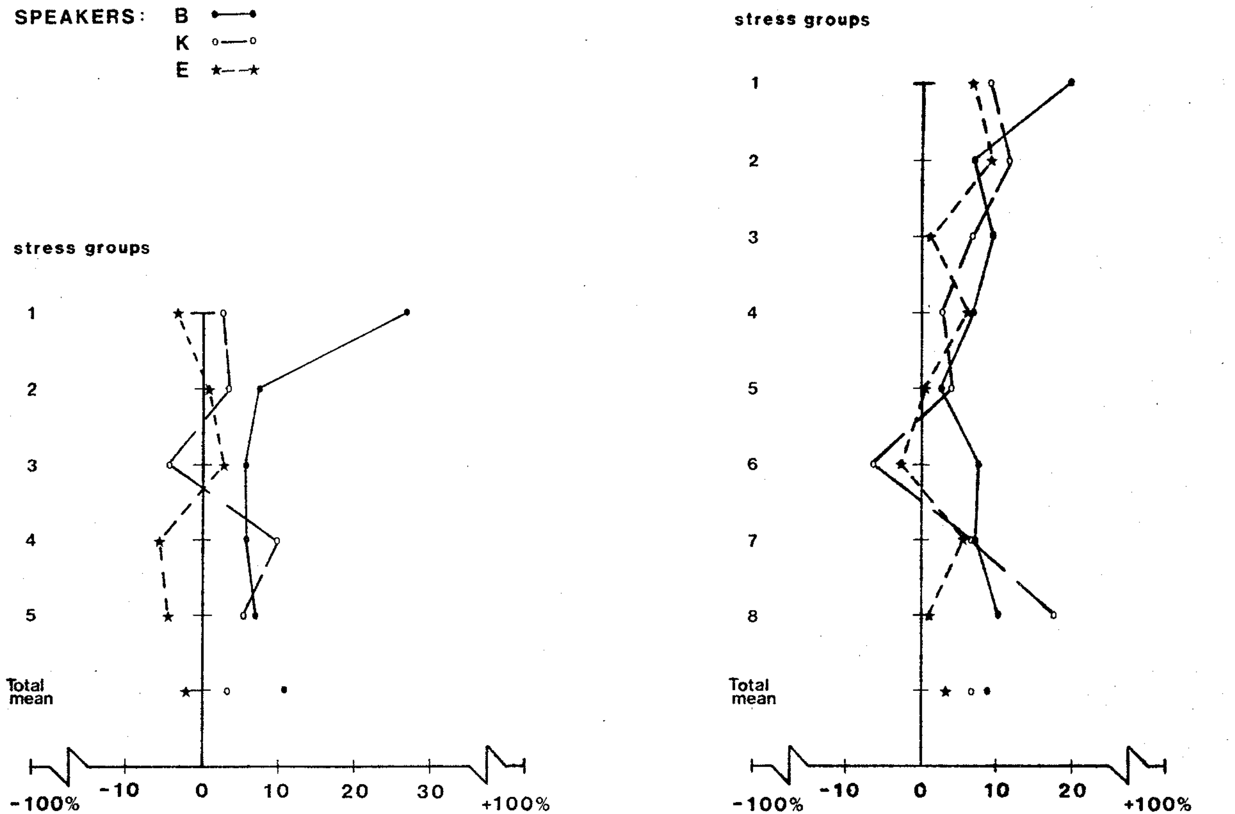

Figure 4. Relative mean differences of stress group durations and total mean in sentences 8 and 14 with 5 and 8 accents (stress groups) respectively between echo question - statement for each speaker, the durations of stress groups of statements being the reference.

Fo. The curves of the three intonation types are superimposed, the VC-boundary of the accentuated syllable being the temporal reference. It is easy to imagine the remaining contours with 4, 5, 6, and 7 accents by a simple operation of interpolation. The horizontal time axis is normalized and determined by the number of accents with equal distances between them. Comparing the different Fo-contours, some characteristic features of German intonation may be summarized as follows: Apart from one exception, each accent is manifested tonally as a rise starting from an Fo-minimum in the initial consonant of the accentuated syllable and ending in an Fo-maximum in the following, unstressed syllable. The only exception to this is the final accent of the statement which is tonally expressed as a fall. As a consequence of this difference in the final part of the utterances, the tonal contrast between the statement with a final fall, signalling terminal intonation and the questions with a final rise, signalling non-terminal intonation is established. The echo question, which also expresses surprise and astonishment, is characterized by at least a somewhat larger tonal range of the tonal movements. The tonal peaks show higher values. Except for the Fo-minimum of the first accent, the bottom points of the accents in all three intonation types show approximately the same value for all speakers. It seems as if the tonal movements of the different utterances started from a common bottom line or as if the tonal contours of the intonation types rested on a common tonal floor.

4.2 The intonation algorithm

In the second step of the basic tonal component of the model,[6] an algorithm generates the tonal structure of a given utterance in five steps. The working of the algorithm in the time-fundamental frequency-field for a sentence with four accents of equal weight, to be spoken as a statement and an information question, is illustrated in Figure 6.

As the first step, the intonation algorithm defines the four basic tonal points or levels which constitute the frame of the whole tonal structure of the utterance. This is done using the given temporal structure as the time reference. In the second step, the Fo-minima of the accents and the end

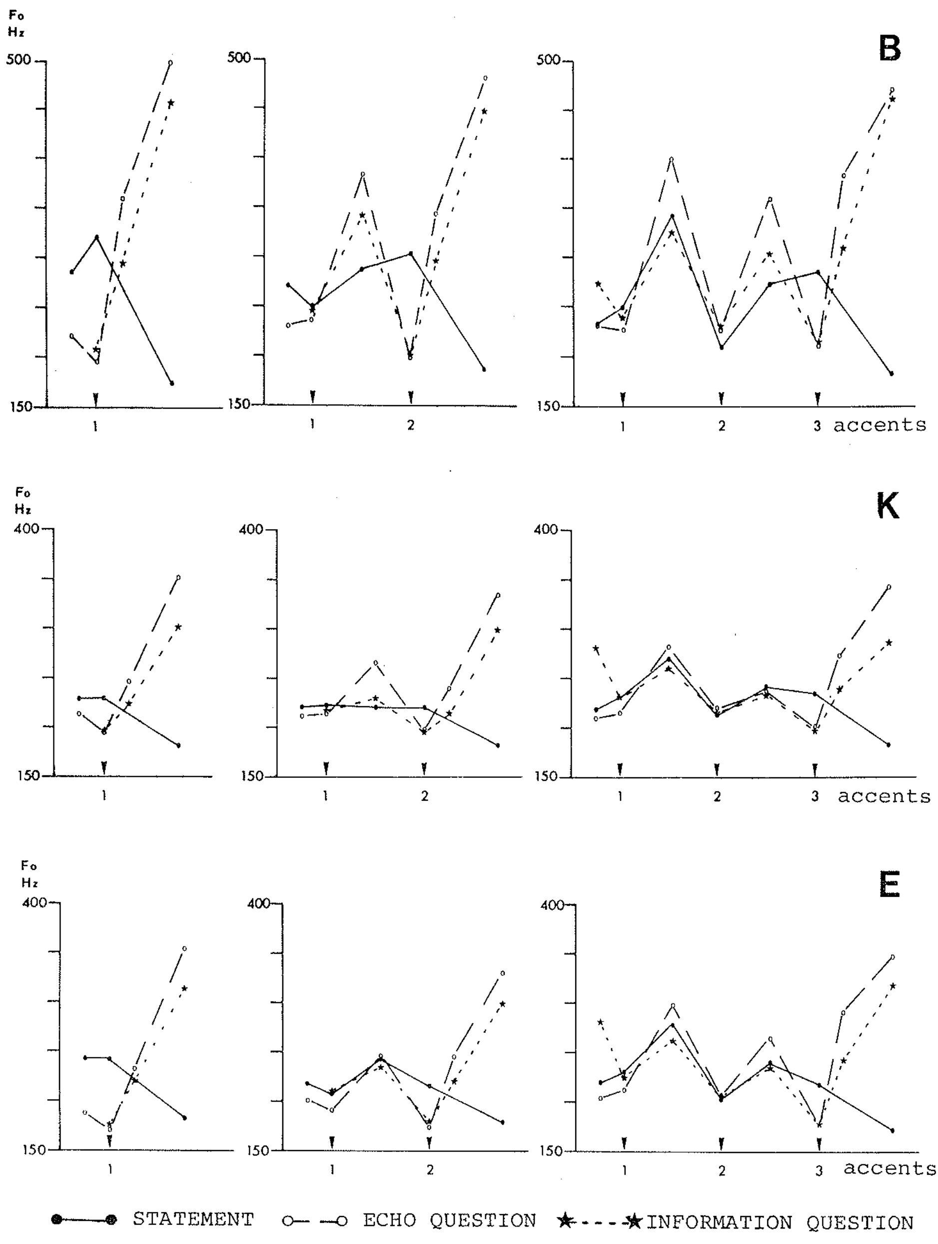

Figure 5. Normalized and superimposed Fo-contours of the three intonation types shown for the sentences with 1,2, and 3 accents (sentences no.1,2,5) for each speaker.

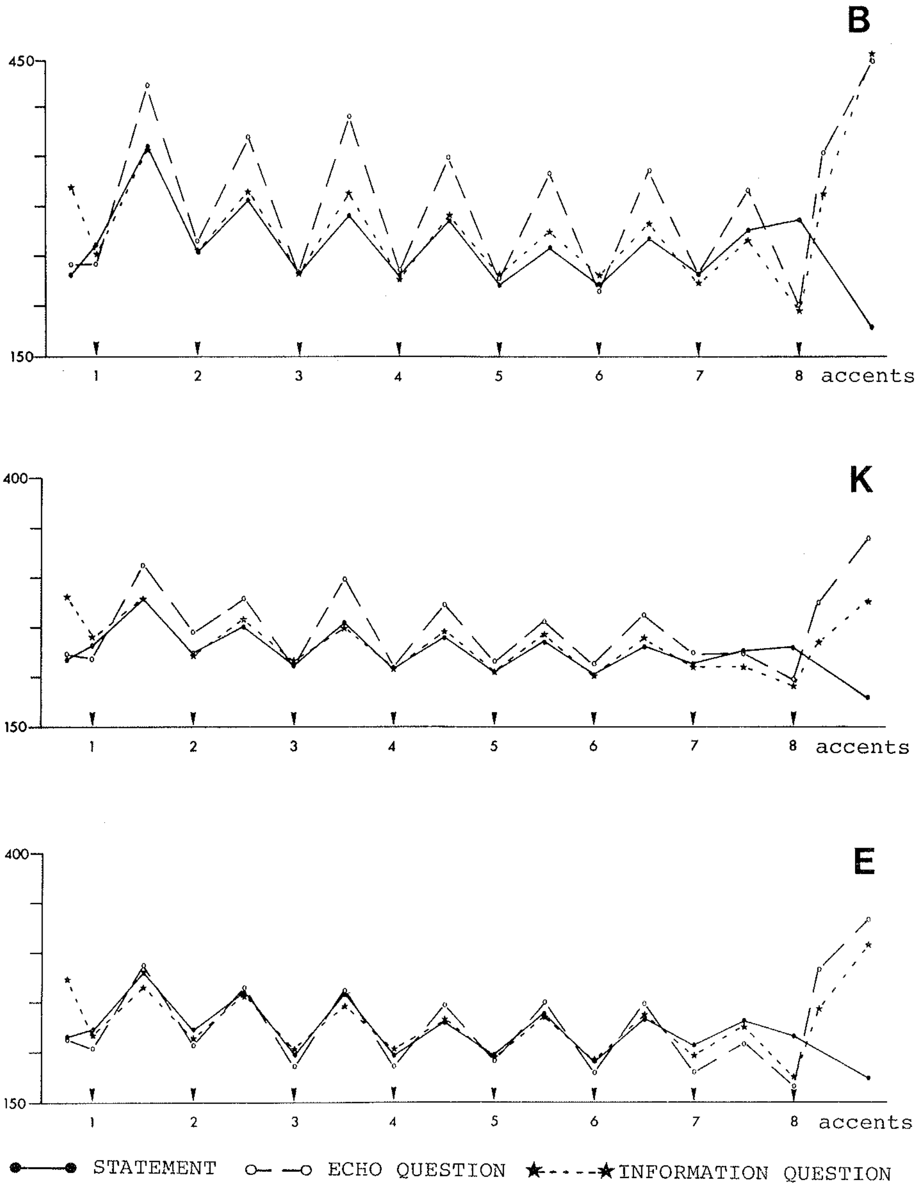

Figure 5. Normalized and superimposed Fo-contours of the three intonation types shown for sentence 14 with 8 accents for each speaker.

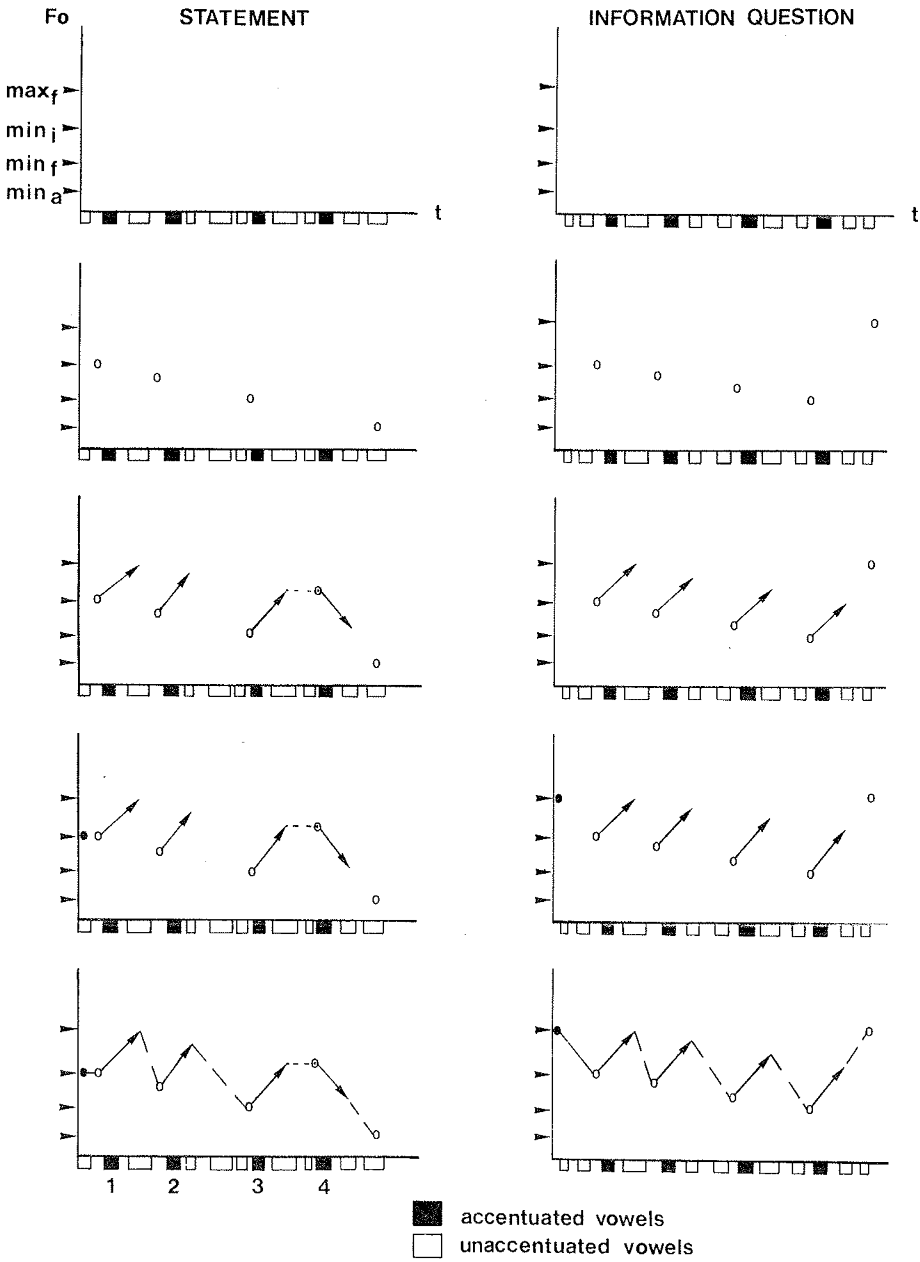

Figure 6. The intonation algorithm illustrating the generation of the tonal structure for a sentence with 4 accents (1,2,3,4) as statement and information question.

point of the utterance (the absolute Fo-minimum of the statement and the final Fo-maximum of the questions respectively) are inserted. Third, the tonal accent movements are generated, starting from each tonal point (Fo-minimum) and thus must be rising by nature. The tonal rises of the accents are of equal size which is defined as a certain tone interval. They end in the vowel of the following unstressed syllable. Therefore, the steepness of the tonal rise may vary according to the temporal conditions. In the final accent of the statement, the tonal accent movement must go down because, as a consequence of the falling contour of the sentence intonation, the final accent of the statement is characterized by a high point. Fourth, the starting points of the Fo-contours are inserted. In the fifth and last step the existing points and fragments of the tonal structure are interconnected either by straight lines or by a cosine function. Now the basic tonal structure of a given utterance is present.

The design of the intonation algorithm clearly shows that the tonal structure is generated from below, i.e. from the tonal bottom or tonal floor which is defined by the Fo-minima. Considering the position of the lowest points of the Fo-contours of the three intonation types (Figure 5), it is quite evident that the corresponding minima of each intonation type have rather similar values. It seems as if the Fo-contours rested in the lowest points, while the peaks varied more freely. Therefore, the main elements of the intonation algorithm are: 1) the Fo-minima of the accents, 2) the final end points of the utterance (low with the statement, high with the questions) which are the most important expressions of the intonation type or the sentence intonation, and 3) the tonal movements of the accents. By utilizing the feature of the tonal accent movements instead of a tonal accent point HIGH, as is the case in the Swedish and American intonation models, the dynamic character of the accents and the tonal changes associated with them is taken into due consideration. The size of the tonal movements varies due to different factors, the most important of which are syntactic, semantic, and pragmatic ones. After all, intonation is no more than kind (timing of tonal changes with reference to temporal units like segments or syllables and thus resulting in a fall or a rise) and size of the tonal changes resting in the tonal floor and their interrelationships.

When constructing the intonation algorithm, several important questions have to be answered:

1. How are the beginning and end of the Fo-contours related to the length of utterances?
2. How do the tonal differences of the accents behave as a function of length of utterance?
3. What does the bottom line, i.e. the distribution of the Fo-minima between starting and end point, look like as a function of length of utterance?
4. What is the tonal expression of the accents and the intonation types? Are the intonation types manifested only locally at the end of an utterance or also globally over the whole phrase?

In the following sections, an attempt will be made to answer these questions. First, the five basic tonal points that define the frame of the whole intonation contour in the first step of the algorithm will be presented and discussed. Secondly, the intermediate tonal points of the floor, i.e. the course of the bottom line, will be investigated and finally, the tonal difference of the accents, the tonal accent rise, will be considered.

4.3 Basic tonal points

The most important tonal points in the intonation algorithm are in a falling scale according to their importance: the Fo-minimum (Min_i) of the first and last accent (Min_f), the end points of the tonal contours, namely the absolute Fo-minimum at the end of statements (Min_a) and the final Fo-maximum (Max_f) of the questions, the starting point of the tonal contour ($Fo_{initial}$), and the peak of the first accent (Max_i). Although this latter point does not belong to the floor of the tonal structure, it may be important for the declination or the global expression of the intonation type (cf. Thorsen 1980, 1981; Bruce 1982) and therefore will be treated together with the other basic tonal points.

When constructing the intonation algorithm, it is essential

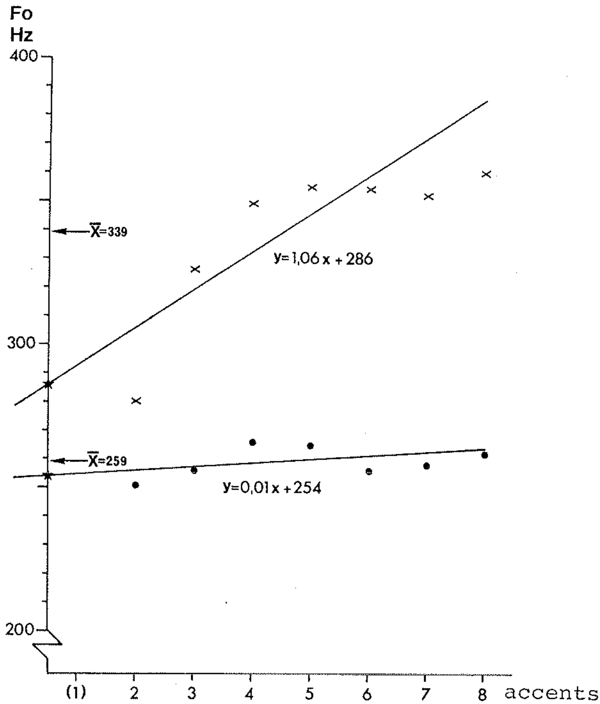

Figure 7. Illustration of the statistical procedure of determining constancy and variation of tonal points.

to know whether or not the most important tonal points, which, as it were, function as the hinges of the tonal structure, are dependent on the factors of sentence length (number of accents) and intonation type. If the points are independent of these two factors, only one value has to be specified in the algorithm. However, if the points are dependent on these two factors (and others), the algorithm has not only to account for the variation per se but has also to determine their kind and size.

In order to motivate the choice of the fixed tonal points of the intonation algorithm, the behaviour of the most important tonal points will be described by statistical measures.
By way of illustration, Figure 7 elucidates the statistical method using the points of the initial Fo-minimum (Min_i) and the initial Fo-maximum (Max_i) of speaker B. Each point in the two series of points represents the mean of the Fo-values for

the sentences containing 2 - 7 accents. The points pertaining to the sentence with only one accent are not taken into consideration. This is due partly to the fact that one of the two points is missing in the contour (cf. Figure 5), and partly to tone assimilation. As a simple statistical measure, the straight line $y = Ax + B$ is fitted to each series of points. Comparing the total mean $\bar{x}$ of one series of tonal points to the intercept B, i.e. calculating the difference $(\bar{x} - B)$, some conclusions concerning the size of the variation of the given tonal point as a function of the utterance length can be drawn. More information is to be found in the slope A of the straight line concerning the size of the variation and in the correlation coefficient r concerning the strength of the relationships.

The one extreme case is illustrated by the points Min_i showing very little variation as a function of utterance length ($\bar{x} - B = 5$ Hz, $A = 0.01$). Thus these points represent tonal constancy. In contrast to this, the other points Max_i vary rather considerably ($\bar{x} - B = 53$ Hz, $A = 1.06$). They represent the case where utterance length exerts a clear effect on the tonal point Max_i, at least up to the length of 4 accents.

The statistical values of the five tonal points for each speaker and each intonation type are presented in Table 3. The following values are shown which are based on the mean for each sentence: The total mean $\bar{x}$ and its standard deviation s in Hz, the intercept B of a straight line fitted to the series of points in Hz, the slope A, the correlation coefficient r, and the difference $(\bar{x} - B)$ in Hz as a measure of the variation of the 7 tonal points as a function of utterance length (number of accents). If it should be the case that a certain tonal point, e.g. Fo-Min_i, has the same value, irrespective of the increasing number of accents and thus utterance length, for instance 200 Hz, then it follows that $\bar{x} = 200$ Hz, $s = 0$ Hz, $B = \bar{x} = 200$ Hz, $A = 0$, $r = 0$, and $(\bar{x} - B) = 0$ Hz. In this case, utterance length has no effect at all on the tonal point Fo-min_i. It has a constant value. This point, however, can not be considered to be constant when $(\bar{x} - B)$ and A increase because then the variation of each point as a function of utterance length also increases.

Table 3. Constancy and variation of five basic tonal points for each speaker and intonation type (see text).

1) $F_o\text{-}Min_i$

	B			K			E		
	A	I	E	A	I	E	A	I	E
$\bar{x}$	259	248	233	229	233	220	224	221	204
s	5.4	4.9	5.5	4.9	6.0	5.5	6.4	6.7	6.6
B	254	245	228	220	220	214	218	220	194
A	0.096	0.05	0.11	0.18	0.25	0.11	0.12	0.03	0.20
r	0.38	0.22	0.42	0.76	0.90	0.45	0.40	0.08	0.66
$\bar{x}$-B	5	3	5	9	13	6	6	1	10

2) $F_o\text{-}Min_{a,f}$

$\bar{x}$	181	204	203	180	197	198	177	178	172
s	3.8	5.2	4.7	1.6	2.6	2.5	2.9	1.9	3.2
B	182	212	203	182	199	196	181	178	176
A	-0.008	-0.18	0.007	-0.048	-0.06	0.05	-0.09	-0.01	-0.09
r	-0.05	-0.86	0.04	-0.73	-0.55	0.51	-0.77	-0.18	-0.69
$\bar{x}$-B	1	8	0	2	2	2	4	0	4

3) $F_o\text{-}Max_f$

$\bar{x}$		454	465		290	341		312	341
s		4.7	15.3		13.1	6.3		8.4	9.9
B		457	488		313	350		312	350
A		-0.07	-0.50		-0.50	-0.19		0.006	-0.21
r		-0.35	-0.80		-0.94	-0.73		0.02	-0.52
$\bar{x}$-B		3	23		23	9		0	9

4) F_o-initial

$\bar{x}$	235	298	237	217	278	215	220	279	207
s	18.7	15.3	5.2	1.2	6.1	4.3	2.4	4.1	3.4
B	237	256	226	218	269	204	225	284	198
A	-0.03	0.76	0.21	-0.02	0.15	0.21	-0.09	-0.10	0.17
r	-0.23	0.93	0.76	-0.32	0.47	0.92	-0.70	-0.47	0.95
$\bar{x}$-B	2	42	11	1	9	11	5	5	9

5) $F_o\text{-}Max_i$

$\bar{x}$	339	342	410	265	266	295	276	264	288
s	28.4	10.8	7.9	20.3	14.9	21.2	12.2	9.9	14.9
B	286	317	396	228	238	249	259	250	266
A	1.06	0.49	0.28	0.74	0.56	0.91	0.35	0.27	0.44
r	0.81	0.98	0.77	0.79	0.81	0.92	0.62	0.59	0.64
$\bar{x}$-B	53	25	14	37	28	46	17	14	22

If a large and systematic variation is observed, the tonal point has to be made flexible in the algorithm.

Table 3 is examined for speaker independent patterns of variation and differences. This means that all three speakers have to show the same behaviour in certain respects. Thus considering the five tonal points and their relationship to the variables utterance length and intonation type, the following statements can be made:

A. Utterance length

Only a slight variation as a consequence of utterance length is to be found with the basic tonal points $Fo\text{-}Max_f$ and $FO_{initial}$, especially with $Fo\text{-}Min_i$ and $Fo\text{-}Min_{a,f}$. Therefore, they are considered to be constant and do not receive a correction factor for utterance length in the model. The slight variation which can be observed in the material and which is quite natural may be achieved in the model not by assigning a fixed value to a given tonal point in each utterance but rather by choosing a varying value. Appropriate values may be chosen from the interval defined as the double standard deviation 2s for each speaker. This choice does not apply to the tonal points only but also to the tonal movements of the accents and the inherent segment durations D_o.

However, there is one point, namely $Fo\text{-}Max_i$, which obviously varies considerably as an effect of utterance length with all speakers. It should be remembered that this variation need not be characteristic of all the utterances with an increasing number of accents (cf. Figure 7). Although the significance of the differences measured in Hz is not clear for perception (it must be examined in listener tests), these differences can be accounted for in the model. However, this will not be done in the intonation algorithm, i.e. in the basic tonal component but rather in the modification component. Utterances with one accent and only a few syllables like sentence 1 (cf. Table 1) have to be modified tonally in different ways because they represent the prosodically most complex cases. All the prosodic features of the sentence have to be manifested in these few syllables and, in a sentence consisting of only one syllable, in just this very syllable.

B. Intonation type

In most cases where the points in the intonation types with regard to the position and to the range are not totally different from each other, the same values are to be taken, irrespective of intonation type.

Identical points are the $Fo\text{-}Min_i$ and $Fo\text{-}Max_i$ of statements and information questions, the $Fo\text{-}Min_f$ of the questions, the $Fo_{initial}$ of statements and echo questions. An obvious difference is to be found with echo questions compared to the two other types, where the points $Fo\text{-}Min_i$ and $Fo\text{-}Max_i$ are concerned. If we assume the enlarged tonal movement of the accents to represent the tonal expression of the echo question, i.e. the tonal range of the movements is increased, the larger value of the $Fo\text{-}Max_i$ is given as the sum (v + w) of the accent rise (cf. Bannert 1983a). This enlarged tonal range of the rise of the echo question, however, is to be found with all accents, not only with the first one. The lower value of the $Fo\text{-}Min_i$ of the echo question compared to the two other intonation types may then be explained as a secondary effect of the enlarged tonal range, assuming the general principle of optimal effect and minimal energy. According to this principle, an optimal tonal difference (the largest tonal range possible) is reached, if both the starting point and the end point of the movement are shifted, which means that the starting point is lowered, while the end point is raised.

A clear difference is to be found between the starting points $Fo_{initial}$ of information questions opposed to statements and echo questions. As this difference results from a different course of the tonal contour at the very beginning, it is accounted for in the algorithm (cf. Figure 6, step 4).

There is furthermore a difference between the high final points ($Fo\text{-}Max_f$) of information questions vs. echo questions with speakers K and E. It should be remembered, however, that the preceding Fo-minima ($Fo\text{-}Min_f$) have the same value. As this difference is not valid for all the speakers, only one point for $Fo\text{-}Max_f$, independent of utterance length and question type, is assumed in the algorithm. However, if it should turn out that the difference observed for Fo-Max_f of the questions would constitute a characteristic and necessary perceptual feature of questions, and thus would represent a linguistic demand, it may be accounted for in the intonation algorithm of the basic tonal component.

4.4 Intermediate tonal points

After having established the beginning and end and the frame of the tonal structure in terms of tonal points (or levels), the question has to be asked how the intermediate points of the tonal floor, i.e. the Fo-minima, should be inserted. In intonation models, a declination or base line is usually constructed which has varying slopes for different intonation types. It should be emphasized, however, that in this outline of a model for German prosody no lines whatsoever are utilized. The reason for this is connected with the view that the tonal structure of an utterance should not be considered as a visual or geometric construction. On the contrary, in this model the dynamic character of the tonal events and their interrelationships are stressed.

As the tonal structure is generated starting from the tonal floor in the intonation algorithm, the position of the Fo-minima forming the tonal bottom is essential. Therefore, their position and distribution between the starting point and end point of the sentences with increasing number of accents and in relation to syntactic phrase boundaries will be shown in some detail.

The position of the Fo-maxima, i.e. the top line, is not considered to be independent and thus a linguistic target per se. The top line is here viewed as a result of the tonal accent movement starting from the tonal floor. Of course, the tonal movements tied to the accents may reflect more prosodic features than just the accents. Other prosodic features concomitant with the accents and superimposed on the tonal movement may be the features CONTRAST, EMPHASIS or PHRASE INTONATION which are expressed in the tonal range feature WIDE and accounted for in the tonal re-writing rules by the factor w (cf. Bannert 1983a). The size fo the tonal movements is determined by the sum of all contributions of tonal features, thus total size of the tonal movement $V_G = (v + w + a + \ldots + n)$.

In Figure 8 the positions of the Fo-minima in the sentences 4 and 5 with 3 accents, 8 och 9 with 5 accents, and 12 and 13 with 7 accents of the three intonation types of speaker B are shown graphically. A very similar picture is to be found with the two other speakers. The pairs of sentences differ with regard to the position of the syntactic phrase boundary which is indicated in Figure 8 (cf. also Table 1). An assumed "declination line", defining the positions of succeeding Fo-minima with equally decreasing frequency values, is drawn for the sake of comparison. This line is defined by the means of the Fo-Min_i and Fo-Min_f of the information question. The reason for choosing these reference points lies in the fact that the final accent has a minimum with questions only and that the point Fo-Min_i of the information question is situated between that of statements and echo questions. In most cases, however, it is closer to the mean Fo-Min_i of statements (cf. Table 3).

In Figure 8 it can be seen clearly that the Fo-minima between the starting points and the end points of the floor in the sentences 5 with 3 accents and 9 with 5 accents occupy positions which show approximately equal frequency differences. This is also true of the sentences with 2 and 4 accents which cannot be shown due to lack of space. In other words, the Fo-minima are distributed equally over the frequency interval between the beginning and the end of the tonal floor. However, the last but one Fo-minimum in longer sentences (with 6, 7, and 8 accents) always has a higher value than the last but two, the highest value is to be found in statements. In spite of this, the final Fo-minimum of the questions (Min_f) and the absolute Fo-minimum (Min_a) of the statements show values similar to those in shorter sentences. It is not so easy to find an explanation for the increased value of the last but one Fo-minimum in the longest sentences. This upstepping of the tonal floor which is shared by all speakers appears independent of the syntactic phrase structure (cf. Table 1). Neither can it be conceived as the tonal assimilation to the Fo-maximum of the final accent of statements because this

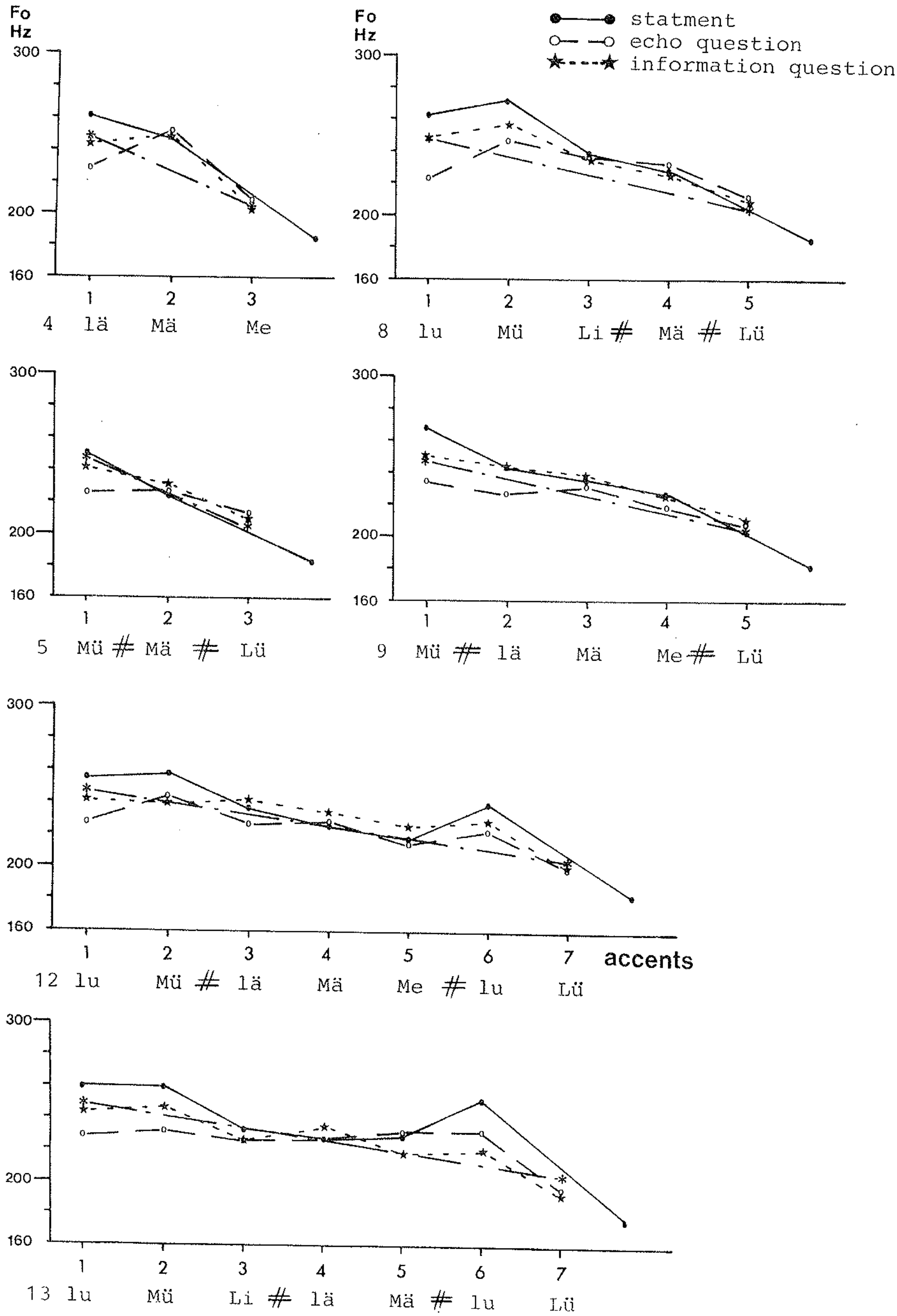

Figure 8. Positions of Fo-minima in sentences with 3,5,and 7 accents and in the three intonation types. Speaker B. A "declination line" is drawn (see text).

upstepping is also found with questions. Tonal assimilation may perhaps account for the higher value in statements compared to questions. One explanation may be presented, though: Due to the large numbers of accents (6, 7, and 8), the frequency intervals of successive Fo-minima become rather small, i.e. the perceptual effect of the successively falling Fo-minima (the downstepping) diminishes, and therefore the speaker has to compensate for this loss of signalling power in order to signal the end of the utterance. As a kind of re-setting, the speaker increases the Fo-minimum of the last but one accent, in order to sharpen the jumping down to the final Fo-minimum of the questions. In statements, the tonal demand of a clearly audible final fall is therefore met by this device.

The successive Fo-minima, however, do not always show diminishing values. The declining course of the tonal floor is broken sometimes. This seems to be the case when there is an effect of syntactic phrase boundaries on the Fo-minima. A comparison between the sentences 5 and 4, 9 and 8 respectively will make this effect clear. While the values of the Fo-minima in sentences 5 and 9, the first phrase of which contains only one accent, decrease gradually, the second Fo-minimum in the sentences 4 and 8 breaks the falling course of the Fo-minima. In both these cases, the first syntactic phrase contains three accents. The value of the Fo-minimum of the second accent is higher than the first value. Thus the second accent in the middle of the three-accent group (Männer in sentence 4 and Müller in sentence 8) appears to be weakened and therefore less prominent than the first and last accent of the group. As an explanation, the principle of tonal grouping of successive accents presents itself. It says that three (or more) successive tonal movements of equal size are not permitted and therefore the second (fourth, etc.) movement will be weakened. This phenomenon might be called the principle of inequality, because it aims at breaking the chain of potentially equal events and thus introduces variation into the tonal contours. This principle may be considered to have its parallel where rhythm is concerned. Rhythmic alternation which means that successive syllables of equal strength are avoided, i.e. one of two equal syllables in succession are weakened, is well known to be characteristic of the rhythmic structure of phrases.

A second explanation may be found on semantic grounds. Due to their meaning, the accents in the syntactic phrases under consideration can not be equal. The place adverbial (Lingen) and the attribute (lullende) which are expressed in the third and first accent respectively, may have a semantically larger weight than the person (Müller) which appears in the second accent. Therefore, the person is moved into the background by using a weakened accent. However, these explanations presuppose that the tonal movement in the second accent, too, is smaller than in the adjacent accent. In the following section (4.5) it will be shown that this is actually the case.

In summary, then, it can be stated that the Fo-minimum of the tonal floor are not always spaced equally between the starting point and the end point of the tonal bottom. In some cases one of the successive Fo-minima is shifted upwards and thus breaks the declining line of the tonal minima. Therefore, it is concluded that various factors, especially semantic features, determine the position of the Fo-minima of the tonal floor.

4.5 Tonal accent movements

The element of tonal movements associated with the accents, which is one of the basic features of the intonation algorithm, expresses the dynamic character of intonation. However, the tonal movements found in the accents, namely rises and falls, do not represent the expression of the accents alone, i.e. the prominence feature of the most important syllable of the lexical units, but also contain reflexes of other prosodic tonal features, like CONTRAST, EMPHASIS, and PHRASE INTONATION which are expressed by the tonal feature WIDE. Finally, the tonal movements are also determined by semantic and pragmatic features.

Figure 9, which is constructed in parallel with Figure 8, shows the tonal movements found in the accents, defined as the tonal difference ΔFo = (Fo-Max - Fo-Min), in the sentence pairs with 3, 5, and 7 accents. The final accent is omitted

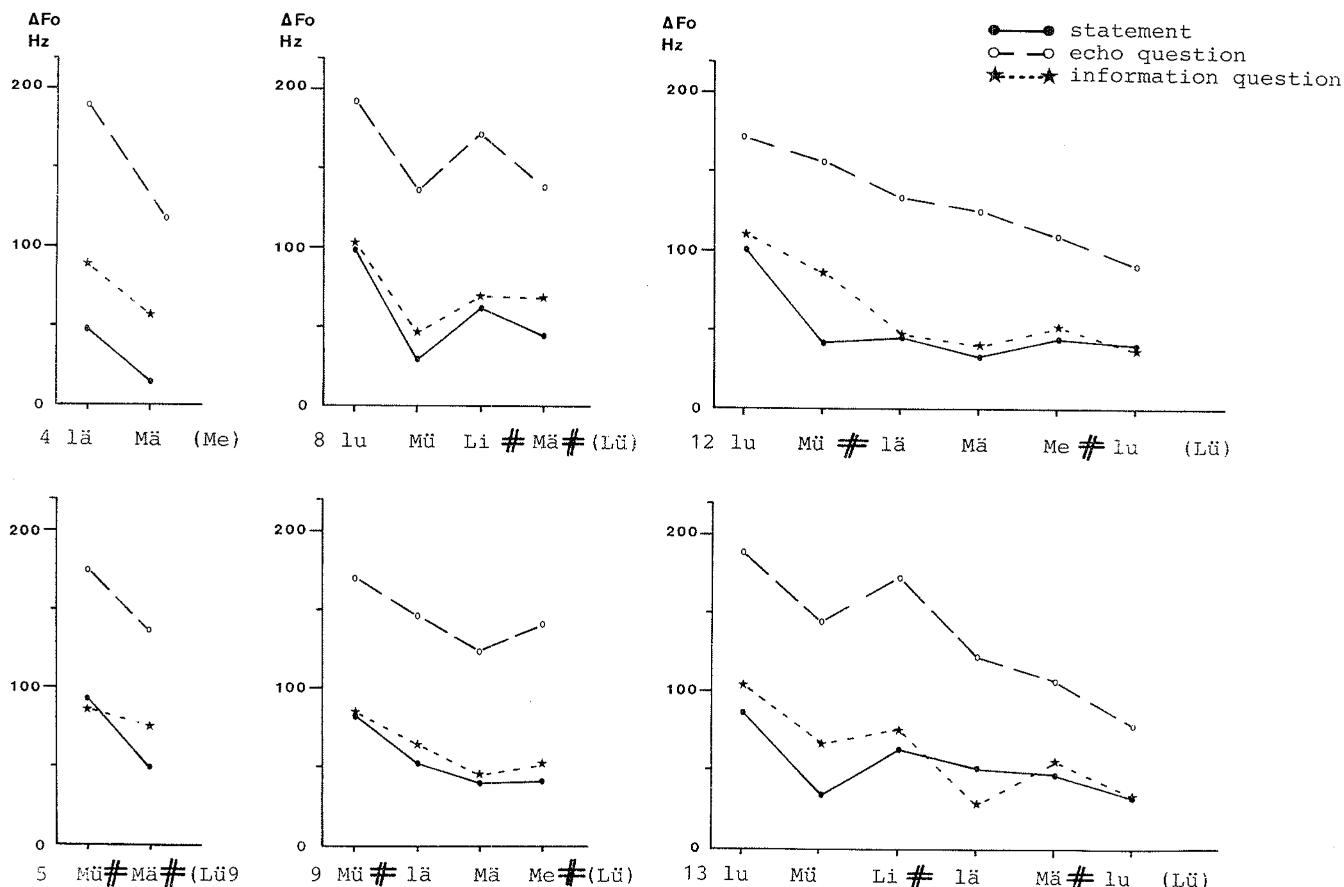

Figure 9. Tonal accent movements in sentences with 3, 5, and 7 accents and in the three intonation types. Speaker B. Sentence numbers, accentuated syllables, and phrase boundaries are indicated.

because it is also affected by the sentence intonation.

The differences of the tonal movements of the accents appear very clearly in the sentences 8 and 9 with 5 accents. The second accent (Müller) of the first syntactic phrase with 3 accents has a clearly smaller tonal rise than the adjacent accents. Except for the statement, this smaller tonal movement is also to be seen in sentences 12 and 13.

Leaving out this irregularity which may be a consequence of the semantic inequality of the three accents of the first syntactic phrase, the tonal movements of the accents, in general, show gradually falling values. In the intonation algorithm, I assume as a rough simplification that the tonal movements of the accents, provided that the accents are of equal weight and the tonal movements do not reflect other tonal features, be given the same value in another scale doing more justice to pitch or the perception of Fo in general. The size of the tonal movements is defined as a constant tone interval. As a consequence of this, the slope of the tonal movement will vary in accordance with the time limitation which is given by the temporal structure. Some evidence for this assumption is to be found in Bannert (1982, 12ff.) where the effect of vowel quantity on the rising Fo-movements of the accents was investigated. The varying slope of the tonal movements is a natural consequence of the definition of the starting point and the end point of the movements. The starting point of the accent, i.e. the Fo-minimum is located in the consonant preceding the accentuated vowel; its end point is located in the following unstressed vowel or the CV-boundary of the post-accentuated syllable.

5. CONCLUSIONS

The acoustic data concerning temporal and tonal aspects of German represent the phonetic basis for the outline of a model for German prosody. It goes without saying, though, that considerable work remains to be done before the model for German prosody will reach the level of sophistication that models for other languages have reached. Some of the most urgent areas of progressing research are linguistic variables of various kinds, the number of speakers and different speech styles, i.e. directed and free spontaneous speech.

Linguistic and phonetic analysis alone, however, will not suffice. The quality of the model for German prosody will heavily depend on investigating how prosodic features of utterances are perceived. As an excellent means for finding out the relevant prosodic features of the signal, given the acoustic description of the temporal and tonal structures of utterances, speech synthesis has to be used. It will be readily agreed upon that it is the aim of our endeavours to understand prosody and its role and function in speech communication as used by both speaker and listener.

FOOTNOTES

1 I would like to point out the preliminary and rather restricted character of this model. Much work remains to be done in order to arrive at a more satisfactory understanding of German prosody. Nevertheless, it can be assumed that some general and language-specific features of German prosody are already captured in this outline.

2 A comparison of the prosody model with those of other languages, a discussion and an assessment of the elements chosen for the model, and other discussions and comments fall outside the scope of this paper. They will need to be saved for future work.

3 The longest sentence 14 reads in English: The miller in Lingen doing pee-pee will always call the taller men in the crowd loitering louts.

4 I am deeply indebted to the director of the Institute, Prof.Dr. Klaus J. Kohler.

5 The story "Bericht über einen geheimnisvollen Zeitungsartikel" (Report on a mysterious newspaper article) was written by Ursel Krützmann and Micheal Weinhold who participated in the seminar "Rhythm and Intonation".

6 Whereas the model for German prosody shown in Figure 1 is also applicable to other languages, the specific features of German intonation are brought out in the intonation algorithm (and the re-writing rules). See also footnotes 1 and 2.

REFERENCES

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

Bannert, R. 1983a. In print.Modellskizze für die deutsche Intonation. Zeitschrift für Literaturwissenschaft und Linguistik 49

Bannert, R. 1983b. Towards a model for German Prosody. Paper presented at the Interdisciplinary Symposium "Prosody - Normal and Abnormal", Zürich, April 6-8, 1983

Bruce, G. 1977. Swedish word accents in sentence perspective. Lund: Gleerup

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

Gårding, E. and Bruce, G. 1981. A Presentation of the Lund Model for Swedish Intonation. Working Papers 21:69-75. Department of Linguistics and Phonetics, Lund University

Thorsen, N. 1980. Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in ASC Danish. Annual Report of the Institute of Phonetics 14:1-29. University of Copenhagen

Thorsen, N. 1981. Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in ASC Danish-Supplementary Data. Annual Report of the Institute of Phonetics 16:19-36. University of Copenhagen