Independence and interdependence of prosodic features

Robert Bannert

ABSTRACT

Considering existing prosody models, two fundamentally different approaches can be discerned. In one case, the tonal structure is treated as independent, whereas the temporal structure is seen as dependent and derivable from the tonal structure. This approach assumes the primacy of intonation. In the other case, the two dimensions of time (duration) and frequency (Fo) are treated as independent of each other. Therefore the one cannot be derived from the other. Instead the basic temporal and tonal structures are generated separately. This approach ascribes time and intonation an autonomous status.

Starting from this dichotomy, this paper will promote the discussion about the principles of building prosody models. It seems essential to abandon the categorical question about the either-or status of time and intonation and to recognize the complex interrelationships between these two dimensions. Therefore time and intonation should be considered equal in principle, although it is quite obvious that there exist certain relationships between them.

An attempt is made to illustrate the approach of equality between duration and Fo using Swedish test material. Aspects of word and sentence level prosody are investigated. The independence and interdependence of duration and Fo will be displayed. The question which is put forward and which seems more fruitful is not *whether* there are any dependencies but rather *what* the interrelationships look like. It will also be demonstrated how tonal features behave in a case of extreme time shortage. When several tonal features of an utterance are forced into one single syllable, a total reorganization of the tonal contour is to be observed exhibiting a clear tonal hierarchy on the word and sentence level.

The observations on the Swedish material are supported by references to equivalent phenomena in some other languages thus lending a more general character to them.

The results of this investigation are the starting point for the outline of a new prosody model. The tonal and temporal structures of utterances will now be generated in parallel with interactive processes. Linguistic rules and information of different kinds are applied. Therefore the adjustment component in an earlier version of the model is disposed of.

Last not least, shedding light on the relationships between time and intonation is important also for the development of high-quality speech synthesis in text-to-speech systems.

INTRODUCTION

For over ten years now, a discussion has been going on concerning the relationship between the two prosodic features of segment duration, i.e. the temporal structure of utterances, and the tonal movements in utterances, i.e. their tonal structure. The question has been whether these two features are independent of each other or if one can be derived from the other. Adherents of the latter view assume that the tonal gestures (movements) constitute the primary, basic feature out of which the segment durations follow as an automatic consequence of the tonal demands and requirements. This stand, which may be termed the primacy of Fo in a prosody and speech model, is taken, for instance, by Öhman et al. (1979) and Lyberg (1981).

Opposing this view, the time and tone dimensions of speech are considered to be separate entities, each of which exists on its own grounds. However, time and frequency do not exist independently of each other. Nevertheless, in a generative prosody model, the basic temporal and tonal structures of an utterance are indeed generated separately of each other. The temporal structure is processed first, because it serves the tonal structure, defined by its tonal anchor points <1>, as a reference for projection. Then the basic temporal and tonal structures are added where different kinds of adjustments become necessary. This is the case when a tonal gesture or successive tonal gestures only have a limited time to be executed. The resulting tonal conflicts are of two kinds: time-dependent and position-dependent (Bruce 1977, 74). The approach which considers time and frequency as separate dimensions, although time is seen as primary delimitating frequency in cases of conflict between them, is represented by Thorsen (1980), Bruce (1977, 1981), Gårding et al. (1982), and Bannert (1982a,b) and may be termed the autonomous model of prosody.

A discussion of the relationships between tonal and temporal features in a prosody model and a first examination of Lyberg's model of Fo-dependent segment duration is to be found in Bannert (1982a).

Taking these opposing approaches as the starting point, it is the aim of this paper <*> to continue the discussion and to arrive at a clearer picture of the principles of a prosody model <2>. It will be asked if it is justified at all to formulate categorical questions about the dependence or independence of time and intonation since data suggest that there is a complex acting together of segment durations and

Fo. Therefore the dimensions of time and frequency should be treated as equal partners and processed separately, although they share independencies and interdependencies. Using Swedish material which contains temporal and tonal variations, these interrelationships will be demonstrated.

Compared to previous studies, the present investigation also widens the number of variables by including the following three variables: (1) the opposite tonal manifestation of identical tonal features (word accent II and sentence accent) in two Swedish dialects (Standard and Southern Swedish), (2) the quantity (complementary length of the stressed vowel and the following consonant in Standard Swedish and long/short vowel contrast in Southern Swedish), and (3) three, different, non-final sentence positions of the test word.

THE INVESTIGATION

The variables are presented that are used for the intended variation of time and frequency. Then the design of the test is shown and information about the recordings and the analysis is given.

Variables

The following variables were changed in a statement spoken as the answer to an appropriate question:

sentence accent
quantity
sentence position of test word (sentence medial)
dialect (Standard Swedish, Southern Swedish)
speakers

For the manifestation of the prosodic features the following differences can be observed:

Besides the tonal differences in the accentuated vowel of word accent II in both dialects (a fall in Standard Swedish, a rise in Southern Swedish), sentence accent is manifested strikingly differently.

There are also dialectal differences as to the manifestation of quantity. Whereas the stressed VC-sequences show the pattern of complementary length (/V:C/ vs /VC:/) in Standard Swedish, Southern Swedish displays quantity in the stressed vowel only (/V:C/ vs /VC /).

Material

As the starting material, the following sentence was chosen which, with respect to its phonetic and syntactic structure, corresponds to a well-established standard in intonation studies of Swedish:

```
Man  kan  l`ämna  l`ånga   n`unnor   efter  `åtta.
          ______  _______  ________
            1        2        3
          verb    adject-  noun
                  ive
(You can leave long nuns after eight o'clock)

` = word accent II (grave accent)
1, 2, 3 = position for test words
```

Test words were <u>stöka</u> with a long vowel and <u>stöcka</u> with a short vowel which were also used in Bannert (1979). The two test words were inserted in turn into the three sentence positions. Sentence accent was placed on the three positions using questions as appropriate contexts. Otherwise, when the test words should not be in focus, sentence accent was placed on the time adverbial (åtta) at the end of the sentence. Thus it was ensured that the word accents were not influenced by the sentence accent because the word accent in position 3 was followed by three unstressed syllables preceding the final word carrying sentence accent. In all, the whole material consisted of twelve sentences: six sentences where the test words did not carry sentence accent, the time adverbial being focussed, and six sentences with sentence accent on each of the three positions and the two test words. Sentence accent was shifted by asking questions about the test words in the different positions (cf. the method used in Bruce 1977, 21 ff.).

Recordings and analysis

The test material was read in a kind of one-person dialogue of question and answer (= test sentence) by four speakers seven times each. Informants were TB (male) and EH (female) from Stockholm (identical with the informants in Bannert

1979) representing Standard Swedish and EK and AO (both female) from Malmö and Lund, respectively, representing Southern Swedish. The sentences were read fluently as one single prosodic phrase, i.e. they were produced in one breath without pausing before the time adverbial. The material was recorded in the acoustic studio of the Department of Linguistics and Phonetics, Lund University, using a STUDER-tape recorder A 62 at the speed of 7.5 ips. The recordings were analysed acoustically using a Frøkjaer-Jensen Pitch Meter yielding a duplex oscillogramme and an Fo curve and, at the same time, a FONEMA Intensity Meter yielding an intensity curve; recording speed was 100 mm/s. The registrations were segmented by hand; segment durations (a[n], s, t, ö, k, a) were measured with an accuracy of 5 ms; Fo was measured at four points A, B, C, D defined below (cf. Fig. 1) with an accuracy of 5 Hz. The individual means were calculated and rounded off to the nearest 1 ms and 1 Hz respectively. Standard deviations were also calculated.

RESULTS

Superimposed tonal contours of a typical utterance containing the test word stöka as an adjective in position 2 with and without sentence accent for both dialects are shown in Fig. 1. The four points of tonal measurement A, B, C, and D are indicated and defined as follows:

Point A: End of the preaccentuated, unstressed vowel.

Point B: Beginning of the accentuated vowel. The Fo-minimum in the Southern Swedish curves are most often preceded by a short, small fall.

Point C: End of the accentuated vowel. The Fo-maximum in the Southern Swedish curves are most often followed by a short, small fall.

Point D: Fo-value at the VC-boundary, i.e. at the end of the unstressed vowel [a] in the second syllable of the test word.

The results are presented as follows:
First, based on the means of all the measured values of segment durations and Fo in the four tonal points, some general observations are made. An overview of the tonal aspects of the material is given in Fig. 2 where the Fo-points are plotted time normalized including all variables for each speaker. Then, based on the mean values, the influence and effect of quantity, sentence accent, and

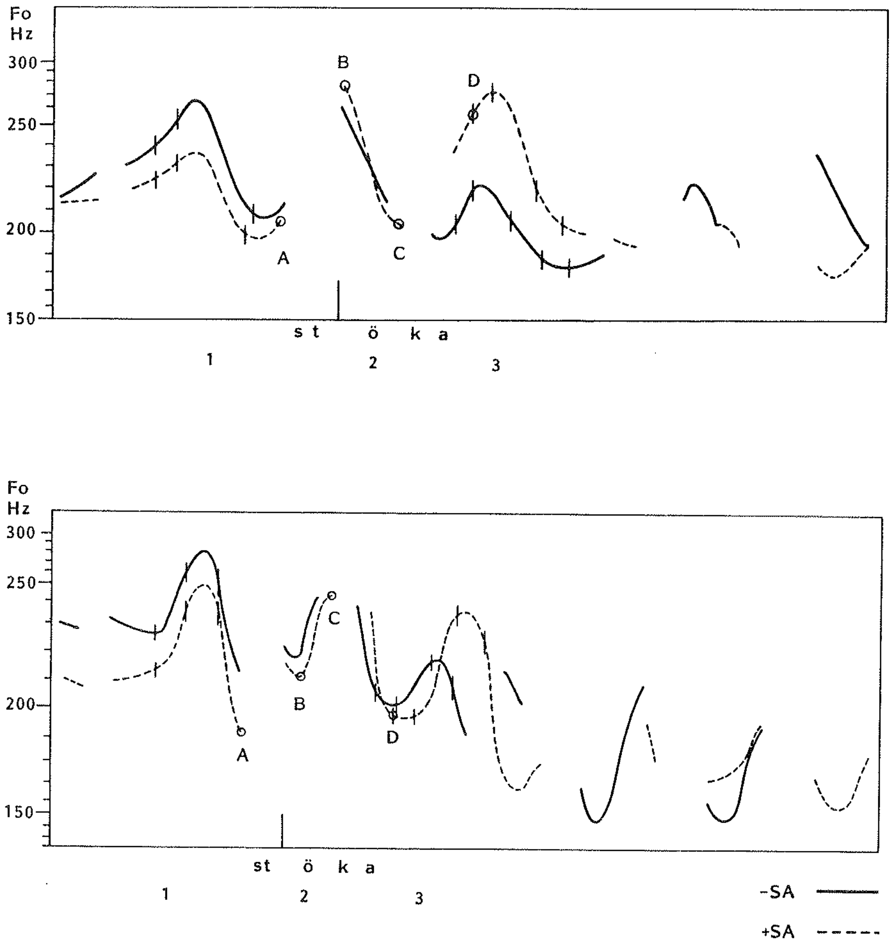

Fig. 1 Superimposed, typical Fo-contours of the test word in medial sentence position (position 2) with and without sentence accent. Four points of tonal reference (A, B, C, D) are shown. Standard Swedish, speaker EH, above; Southern Swedish, speaker EK, below.

sentence position on the temporal and tonal structure of the test words respectively are reported. In each case, the differences calculated from the means are given in tables. Finally a conflict situation between time and frequency is shown where time dominates over frequency, and a tonal hierarchy at work is illustrated.

Segment durations and Fo-values

On the basis of mean values, the following general observations can be made:

Duration

1. Sentence accent increases the duration of all segments with all four speakers, although in some cases only to a small extent. There are also instances, however, where the increase is considerable. There is only one exception: Speaker AO, final [a], all positions, where a systematic decrease of segment duration is to be found.

2. Many segments show smaller durations in sentence position 2 compared to the other positions. This seems to be a positional effect, i.e. an expression of the rhythmical organization. According to this principle, a succession of two or more equal accents is avoided by weakening the accent in the middle temporally as well as tonally (cf. Bruce 1983 for Swedish and Bannert 1983 for German).

3. The VC-sequences of the Stockholm speakers show the typical pattern of complementary length (cf. for instance Elert 1964) which the Southern Swedish speakers do not have. Their consonants following short vowels are only slightly longer (cf. Gårding et al. 1974).

Fo

Fig. 2 clearly shows the different tonal behaviour of the Fo-points and the Fo-movements with and without sentence accent, with long and short vowels, in the three sentence positions, between the two dialects, and between the two speakers of each dialect. It should be noted, however, when inspecting the curves that only the movement between point B and C (beginning and end of the accentuated vowel) is to be seen completely in the registrations (cf. Fig. 1). The other two points are simply connected by straight lines. For the

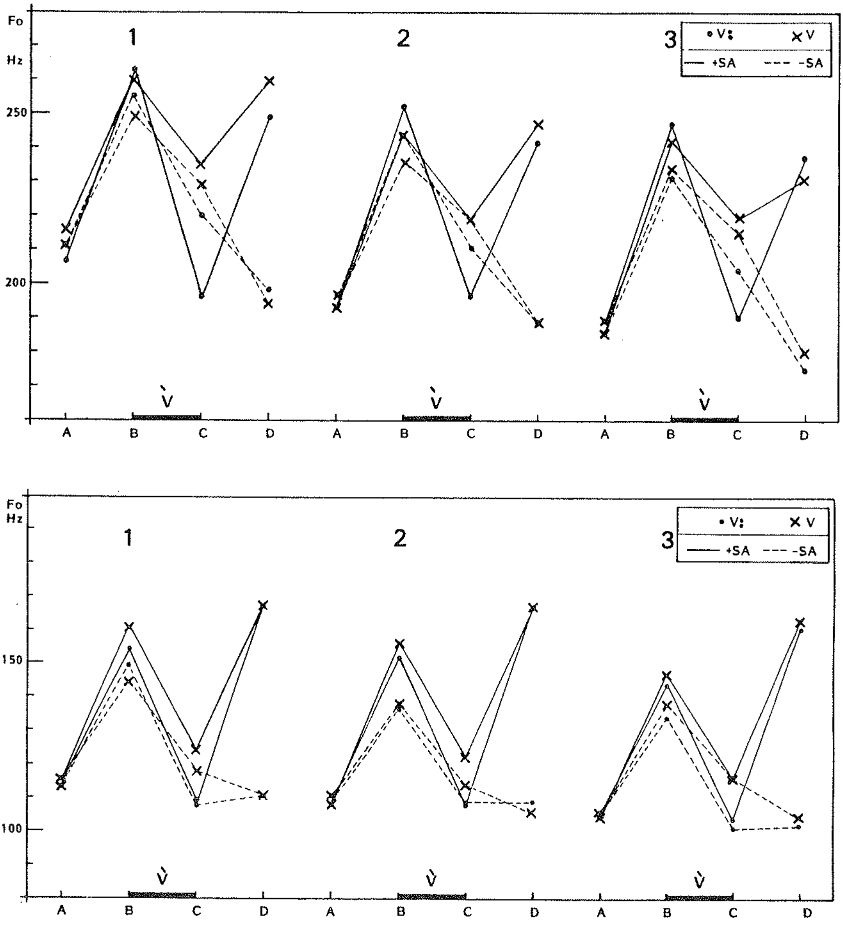

Fig. 2a. Superimposed Fo-contours (time normalized) of the test word in the three positions 1, 2, 3 for the Standard Swedish speakers (EH above, TB below). Four conditions: Long/short vowel, with and without sentence accent.

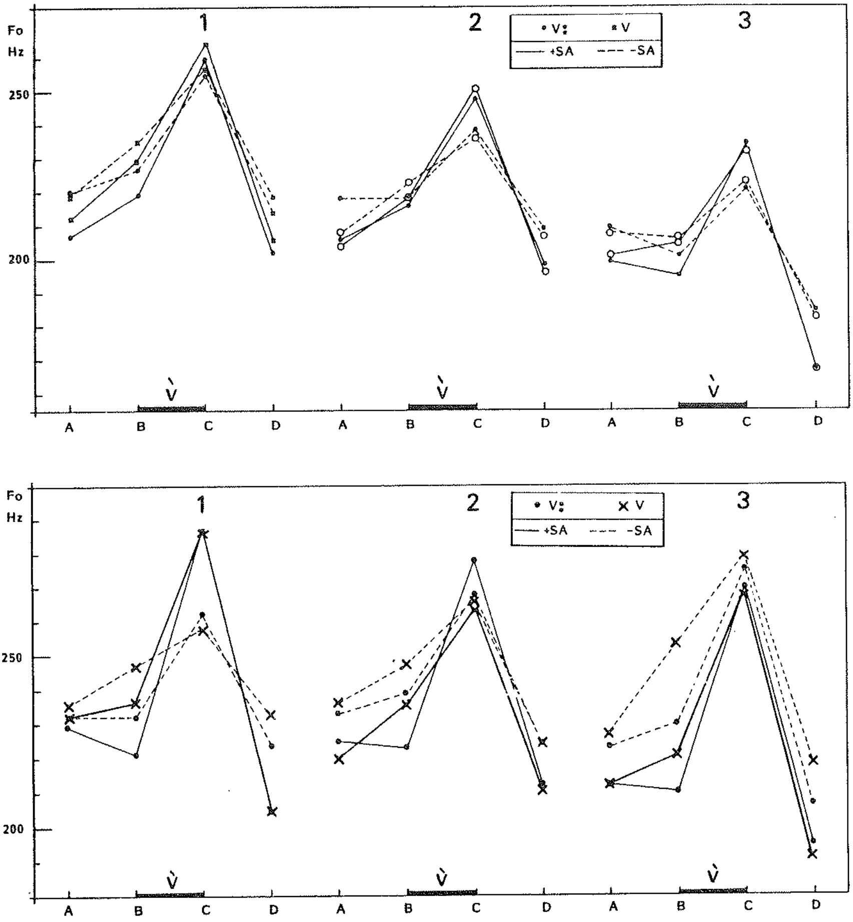

Fig. 2b. Superimposed Fo-contours (time normalized) of the test word in the three positions 1, 2, 3 for the Southern Swedish speakers (EK above, AO below). Four conditions: Long/short vowel, with and without sentence accent.

aim of this investigation, the variation of the Fo-points is of interest and not the complete Fo-movement.

All variables show an effect on the Fo-values, although the shape and movement of the tonal contours, by and large, are preserved in each case. The tonal gesture of the word accent is treated differently in the two dialects. Whereas the tonal fall of accent II in Standard Swedish is truncated, the tonal rise in Southern Swedish is reorganized (cf. Bannert and Bredvad-Jensen 1975). Sentence accent shows up tonally not only in the post-accentuated syllable - as a high point in Standard Swedish and as a low point in Southern Swedish - but also in the accentuated syllable itself (this is clearly to be seen in the curves of the Southern Swedish speakers). Thus sentence accent exerts an influence tonally and temporally on the whole test word. This influence, though, is still greater where the pre-focal accent and the overall shape of the sentence contour is concerned (this effect can be clearly seen in Fig. 1).

Quantity and sentence position also affect the tonal structure. In Fig. 2, the Fo-declination throughout the utterance is to be seen exhibiting differnces between the speakers.

What, then, are the effects in this particular case that quantity, sentence accent and sentence position have on the tonal and temporal structure of the test word in both dialects? We will look for patterns of variation or consistency that can be found either in the whole material or in one dialect, respectively.

Quantity

Durations

The durational changes of all test segments as a consequence of quantity are calculated and given in Table 1. A minus sign indicates that the duration of a given segment is larger in the word with the short vowel. As for the following tables, the values in Table 1 are derived from the means of the basic data. For the sake of simplicity, the standard deviations are omitted <3>.

Table 1 shows that, as a rule, the duration of all segments varies in all conditions. The largest durational differences

Table 1. Differences of segment durations (ms) due to quantity.

POSITION						1								2								3				
SEGMENTS			an	s	t	ö	k	a	VC	WORD	a	s	t	ö	k	a	VC	WORD	a	s	t	ö	k	a	VC	WORD
STOCKHOLM																										
	TB	+SA	2	5	2	79	-41	0	38	43	1	1	0	79	-54	15	25	26	0	1	-3	75	-55	-7	20	10
		-SA	-4	-3	3	33	-21	-5	12	4	3	10	0	30	-18	-7	12	10	5	2	-5	43	-18	-2	25	13
	EH	+SA	-5	6	-5	50	-46	-14	4	-7	1	0	8	50	-34	3	16	27	3	9	5	63	-39	-9	24	29
		-SA	1	8	-4	29	-20	-11	9	0	-1	1	4	25	-24	-3	1	16	2	1	2	36	-21	-5	15	8
SKÅNE																										
	EK	+SA	-1	-1	-10	34	-15	-3	19	5	-3	-8	-8	32	-14	-4	18	-2	1	-7	-5	42	-15	3	27	15
		-SA	2	-1	-7	35	-11	-7	24	11	-4	-8	-2	28	-9	-6	19	3	-2	-2	-10	35	-13	-10	22	1
	AO	+SA	9	0	-13	57	-38	-2	19	2	4	-2	-10	61	-12	-6	49	33	3	1	-10	65	-21	3	44	37
		-SA	6	15	-5	44	-5	-3	39	42	-1	8	-11	47	-9	-5	38	30	2	-4	-2	45	-12	-9	33	13

Negative values indicate that the segment durations with the short vowel are larger than with the long vowel.

are to be found in the accentuated vowel and the following consonant, especially in the VC-sequence with sentence accent. Most clearly, this difference is to be seen with the Stockholm speakers. Thus the VC-sequences display a consistent pattern of variation. The other segment durations vary, by and large, only to a small extent and without any discernible pattern.

Fo-points

Table 2 shows the differences of the Fo-means in the four tonal points A, B, C, and D as a consequence of quantity (cf. Fig. 2). The minus sign indicates that the Fo-value in the test word containing the short vowel is greater (the point is higher) than in the test word with the long vowel (cf. Fig. 1). In this case, except for a few instances especially in sentence position 2, a consistent pattern of variation for the whole material is to be found. The Stockholm speakers show a tonal difference at the end of the vowel (Fo-point C) which is considerably larger than at the beginning. The end point of the short vowel contour with and without sentence accent in each position is clearly higher than that of the long vowel contour. The Southern Swedish speakers show the largest tonal difference at the beginning of the vowel (Fo-point B), the short vowel causing the highest Fo-values. The Fo-values in the other points A and D, in general, vary only slightly and inconsistently.

Sentence accent

Durations

With a few exceptions, sentence accent causes an increase in the segment durations. Table 3 gives the differences of segment durations as a consequence of sentence accent. The minus sign indicates that the segment duration in the test word without sentence accent is larger than in the test word with sentence accent. Large and systematic variations are to be found in the segments [s], the accentuated vowel [o], the following consonant [k], and the unstressed vowel [a]. The smallest and most inconsistent durational changes are to be observed in the segments [a(n)] and [t]. In Standard Swedish,

Table 2. Differences of Fo (Hz) at the four Fo-points due to quantity.

POSITION			1				2				3			
FO-POINTS			A	B	C	D	A	B	C	D	A	B	C	D
STOCKHOLM		+SA	-1	-7	-15	0	1	-4	-14	1	1	-3	-12	-3
	TB	-SA	-3	5	-10	0	-1	-1	-5	3	0	-4	-15	-2
		+SA	-8	3	-37	-10	1	8	-23	-5	1	5	-29	6
	EH	-SA	0	6	-9	4	1	8	-8	0	-1	-3	-11	-4
SKANE		+SA	-5	-11	-5	-4	2	-2	-3	2	-2	-10	1	0
	EK	-SA	1	-8	-2	5	10	-5	2	2	1	-5	-2	2
		+SA	-3	-15	1	0	5	-13	14	3	0	-11	2	4
	AO	-SA	-3	-15	4	-8	-3	-8	2	0	-4	-23	-4	-10

Negative values indicate that the Fo-points with short vowels are higher than with long vowels.

Table 4. Differences of Fo (Hz) at the four Fo-points due to sentence accent.

POSITION			1				2				3			
FO-POINTS			A	B	C	D	A	B	C	D	A	B	C	D
STOCKHOLM		V:	1	5	1	57	0	15	-1	58	1	10	7	59
	TB	V	0	17	6	57	-2	18	8	61	0	9	0	60
		V:	-4	8	-24	51	-2	12	-15	53	3	16	-14	62
	EH	V	4	11	6	65	-2	8	0	58	-3	8	4	52
SKANE		V:	-13	-8	5	-17	-12	-2	10	-11	-10	-6	13	-18
	EK	V	-7	-5	12	-8	-4	-5	15	-11	-7	-1	10	-16
		V:	-3	-11	26	-19	-8	-16	10	-12	-11	-20	-5	-12
	AO	V	-3	-11	29	-27	-15	-11	-2	-15	-15	-32	-11	-26

Negative values indicate that the Fo-points with sentence accent is lower than that without sentence accent.

Table 3. Differences of segment durations (ms) due to sentence accent.

POSITION					1								2								3				
SEGMENTS		an	s	t	ö	k	a	VC	WORD	a	s	t	ö	k	a	VC	WORD	a	s	t	ö	k	a	VC	WORD
STOCKHOLM																									
TB	V:	5	22	15	56	27	23	83	143	1	16	12	61	26	19	87	130	1	16	9	43	31	26	74	135
	V	-1	14	16	10	47	16	57	104	3	25	12	12	62	7	74	114	6	17	7	11	68	31	79	138
EH	V:	8	18	-1	35	16	8	51	76	2	19	4	43	20	11	63	96	8	19	3	44	17	13	61	96
	V	12	20	0	14	42	11	56	83	0	20	0	17	30	5	47	73	5	11	0	17	35	19	52	77
SKÅNE																									
EK	V:	-1	11	2	13	9	5	22	36	-7	9	-3	18	7	2	25	32	2	16	3	23	18	42	41	101
	V	2	11	5	14	13	1	27	43	-8	9	3	14	12	0	26	37	-1	21	-2	16	20	29	36	87
AO	V:	2	11	1	27	31	-1	58	73	-3	-3	5	17	15	-3	32	30	1	19	1	27	25	-7	52	69
	V	-1	26	9	14	64	-2	78	113	-8	7	4	3	18	-2	21	26	0	14	9	7	34	-19	41	45

Negative values indicate that the segment durations without sentence accent are larger than those with sentence accent.

the durational differences are largest in the long segment (V: and C:), respectively.

Fo-points

Sentence accent also affects the Fo-values. Table 4 shows the Fo-differences in the four tonal points A, B, C, and D as a consequence of sentence accent. The minus sign indicates that the value of the Fo-point is larger in the test word without sentence accent, i.e. it is higher. No consistent pattern for the whole material can be found. Within the two dialects, however, there is a similar tonal behaviour.

With the Stockholm speakers, the large tonal difference appears in point D. At this point (VC-boundary), the Fo-maximum of the sentence accent is almost reached. The high tonal point at the beginning of the word accent fall (point B) is also higher with the sentence accent, both the long and short vowel and in each sentence position. Point A, the Fo-minimum in the pre-accentuated syllable remains nearly unchanged. At point C, the Fo-minimum of the word accent fall, the speakers behave differently. Whereas speaker TB hardly varies in this point, speaker EH makes the word accent fall with sentence accent end considerably lower only in the long vowel.

The picture is more uniform with the Southern Swedish speakers. The Fo-points A, B, and D are lower with sentence accent, point C is higher (one exception: speaker AO, position 2, short vowel and position 3). This means that the tonal movement before and in the test word is larger with sentence accent, i.e. the Fo-curve makes a larger excursion up and down, the tonal movement shows a larger range (cf. Fig. 2).

Positions

Durations

Segment durations vary also as a consequence of sentence position. Table 5 gives the durational differences between the positions where the value of position 2 serves as a reference. The first value in Table 5 corresponds to the

Table 5. Differences of segment durations (ms) between positions.

SEGMENTS				a(n)	s	t	ö	k	a	VC	WORD
STOCKHOLM											
	TB	V:	+SA	18	1	6	0	3	-9	3	5
				5	1	-6	0	9	8	9	16
			-SA	14	-5	3	5	2	-5	7	-8
				5	1	-3	18	4	1	22	11
		V	+SA	17	-3	4	0	-10	-4	-10	-11
				6	1	-3	4	10	20	14	32
			-SA	21	8	0	2	5	-7	7	-2
				3	9	2	5	4	-4	9	8
	EH	V:	+SA	12	18	-11	8	1	-13	9	0
				6	11	-1	16	4	-3	20	28
			-SA	6	19	-6	15	5	-10	20	20
				0	11	0	14	7	-5	21	28
		V	+SA	18	12	2	8	13	4	21	34
				4	2	2	3	9	11	12	26
			-SA	4	12	2	11	1	-2	12	24
				-3	11	2	3	4	-3	7	22
SKÅNE	EK	V:	+SA	-2	6	-1	9	0	3	9	16
				12	15	4	20	24	49	44	111
			-SA	-8	4	-6	14	-2	0	12	11
				3	8	-2	15	13	9	28	44
		V	+SA	-4	-1	1	7	1	2	8	9
				8	14	1	10	25	44	35	94
			-SA	-14	-3	-1	7	0	1	7	3
				1	2	6	8	17	13	25	44
	AO	V:	+SA	14	4	-10	6	8	2	14	22
				14	3	-9	12	20	13	32	54
			-SA	9	2	-6	-4	-8	0	-12	-20
				10	-5	-5	2	10	17	12	16
		V	+SA	9	14	-7	10	34	-2	44	53
				15	14	-9	8	29	4	37	50
			-SA	2	-5	-12	-1	-12	-2	-13	-33
				7	7	-14	4	13	21	17	33

Line above: difference between 1st and 2nd positions, negative value indicates that the segment duration in the 1st position is smaller than that in the 2nd position

Line below: difference between 2nd and 3rd position, negative value indicates that the segment duration in the 2nd position is smaller than that in the 3rd position durational difference of a given segment between position 1 and 2, the second value to that between position 2 and 3. A negative value indicates that the segment duration in the first position is smaller than that in the second position and that the segment duration in the second position is smaller than that in the third position <4>. As Table 5 shows, no pattern of variation of segment duration between positions can be found, neither for the whole material, nor for each dialect, nor for each speaker. This is also true of the duration of the VC-sequence and the word.

Fo-points

With only a few exceptions, above all for speaker AO, the position of the test word in the sentence, i.e. its placement in the tonal contour of the sentence with reference to the time axis, affects the Fo-points in a systematic way. Table 6 gives the differences in Hz of the four tonal points A, B, C, and D between the sentence positions (cf. Fig. 2). The first value in Table 6 is the Fo-difference between positions 1 and 2, the second value is the Fo-difference between positions 2 and 3. A minus sign preceding the first value indicates that the Fo-value in position 2 is larger than that in position 1. A minus sign preceding the second value means that the Fo-value in position 3 is larger than that in position 2. Table 6 shows that each Fo-point decreases the further it is located to the right in an utterance. Thus all the four points obey the following rank order: 1 < 2 < 3, i.e. the Fo-values are largest in position 1 and smallest in position 3. In other respects, however, no systematic variation is to be observed. Nevertheless, there are individual differences as to the size of the tonal differences according to position. Whereas the Fo-values of speaker TB only show small differences, they drop, sometimes considerably, with the other speakers. This means that the Fo-declination of speaker TB is rather small, the other speakers showing a clear declination (cf. also Fig. 2). Speaker AO manifests the largest irregularities in her tonal variation.

The two-dimensional analysis of this investigation has shown that the tonal and temporal features affect each other mutually to some degree. However, the prosodic features themselves are preserved as characteristic temporal and tonal patterns. Sentence accent turns out to be a prosodic feature with a Janus face; it is clearly signalled tonally as well as temporally over the whole word which it makes prominent on

Table 6. Differences of Fo-points (Hz) between positions.

FO-POINTS				A		B		C		D	
DIFFERENCE BETWEEN POSITIONS				1-2	2-3	1-2	2-3	1-2	2-3	1-2	2-3
STOCKHOLM											
	TB	V:	+SA	5	3	2	8	1	4	0	7
			-SA	4	4	12	3	-1	8	2	7
		V	+SA	7	3	5	9	2	6	1	3
			-SA	5	5	6	0	4	-2	5	2
	EH	V:	+SA	14	4	11	5	0	6	7	5
			-SA	16	9	11	13	9	7	9	14
		V	+SA	23	6	16	2	16	0	12	16
			-SA	17	5	13	2	10	4	5	10
SKANE											
	EK	V:	+SA	1	7	3	21	12	14	4	31
			-SA	2	9	9	17	17	17	10	24
		V	+SA	8	3	12	13	14	18	10	29
			-SA	11	0	12	17	19	13	7	24
	AO	V:	+SA	4	13	-2	13	10	8	-7	19
			-SA	-1	10	-7	9	-6	-7	0	19
		V	+SA	12	8	0	15	23	-4	-4	20
			-SA	0	8	0	-6	-8	-13	8	9

the sentence level.

In the present investigation, the tonal features in each position had enough time (syllables) at their disposal in order to be manifested without difficulty. However, a positive statement about the independence or dependence of variables can best be made when a conflict arises between the dimensions. Such a case of conflict between tonal features and the temporal structure of the utterance, i.e. between frequency and time, will be demonstrated in the following section.

A tonal hierarchy

Starting from a phrase with four syllables and the three tonal features of word accent I, phrase (sentence) accent and terminal juncture (statement), a stepwise reduction of the number of syllables will show the dependence of tonal features on time. The tonal feature, with the largest domain will dominate over the other features which rank lower in the tonal hierarchy.

Consider the four Swedish phrases:

å 'länderna	"and the countries"
å 'länder	"and countries"
å 'land	"and country"
'land	"country"

As the number of syllables is decreased, the duration of the phrase is also decreased, namely from about 750 ms to about 400 ms. The number of the tonal features, however, is kept constant. Thus the one-syllable utterance is produced with the same tonal features as the four-syllable utterance. Fig. 3 shows the changes of time and Fo-contour in each utterance spoken by a Stockholm speaker. In this time-dependent case of conflict, it becomes obvious that the complex overall Fo-contour shrinks considerably. The tonal contour is reorganized displaying a clear hierarchy between the three tonal gestures each of which can be seen in the complete tonal contour (cf. also Bruce 1977, 92 ff.). The final fall associated with the terminal juncture is preserved in the one syllable phrase but, at the same time, it becomes steeper and "moves" to the left on the time axis into the final (and only) accentuated vowel. The tonal movements associated with the two other gestures are skipped. The subordination of

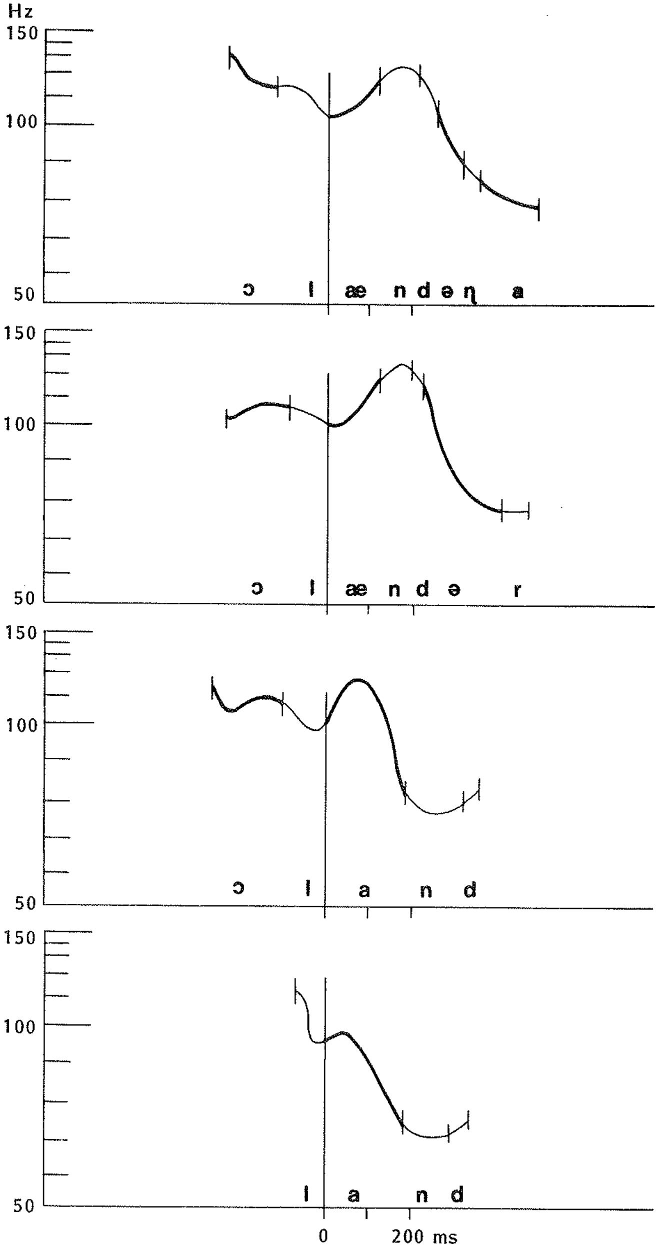

Fig. 3 Hierarchical order of tonal features illustrated by a step-by-step shortening of the duration available for three tonal features. The final fall of the terminal juncture dominates over the features of phrase accent and word accent which have smaller domains.

tonal features of a temporally smaller domain under the feature of terminal juncture with the largest domain is to be observed in other languages too. Even if the hierarchically lower gestures are extinguished from the Fo-curve, they, however, do not disappear for the listener. They will be reconstructed and thus "heard" as a consequence of the remaining tonal movement of the hierarchically highest feature drawing upon the knowledge of the rules of tonal variation in Swedish with reference to the syllable structure of the utterance and thus, finally, to its duration.

DISCUSSION

Before outlining a new concept of a prosody model, some aspects of the interrelationships between time and Fo will be discussed.

Interrelationships between time and Fo

The results of the present investigation show that there is no simple relationship between the temporal and tonal structure of an utterance. Instead both structures are connected and interlocked with each other in different ways. Segment durations and Fo-movements that are to be found in the speech signal result partly from autonomous prosodic features of time and tone, partly from their mutual effects (tonal-to-temporal and temporal-to-tonal), and partly from intervening factors like speech tempo and the individual behaviour of the speaker. Thus, in conclusion, the relationships between duration and Fo in speech are not simple and unidirectional, but rather complex and bidirectional.

After all, is it really justified to ask the categorical question as to the independence of time and Fo-structure? A question which leads, for instance, to the view of the primacy of Fo. As a matter of fact, the tonal gestures or features, alone or in combination, are assigned to certain linguistic units like vowel, syllable, stress group, phrase, sentence, and text. Therefore the prosodic gestures are associated phonologically with linguistic units in some arrangement which is to be thought of as linear and punctual. However, even abstract tonal gestures are related to time because these linguistic units have to be projected onto the time dimension when manifested.

Time remains an autonomous dimension of prosodic features, even if we, as the gesture theory (Öhman et al. 1979, Engstrand 1983), assume that all the phonological gestures, spectral and prosodic, are not defined in temporal terms, but appear, for a given utterance, in an abstract string of simple or combined gestures and which are coarticulated together. When all these timeless gestures are executed, all of them, nevertheless, cannot come out as a natural consequence of their essential conditions and requirements. Some phonological gestures are, sui generis, temporal by nature. One and the same gesture may result in very different segment durations, according to context. The s-gesture, for instance, will be executed temporally in different ways depending on the segmental and prosodic context. Second, segment duration also varies in passages, context being constant, where Fo does not change and thus the constant Fo does not put any requirements at all on a given spectral gesture. Third, the view is generally accepted that the length distinction of quantity in Swedish and other languages, such as Danish and German, is tied to stress, i.e. quantity can only appear in stressed syllables. In this respect, a coarticulation of prosodic features exists. In unstressed positions, however, a reduction or neutralisation of quantity takes place, as is often the case with spectral gestures, for instance vowel reduction in English and Russian. In other quantity languages, like for instance Finnish and Czech, the quantity gestures also appear in unstressed syllables. Thus they are independent of stress in these languages.

The autonomy of temporal patterns will be even more obvious in a contrastive perspective. Take for instance the gesture or gestures for a voiceless, word-medial /p/ which we assume to be identical in languages, such as Finnish, Standard Swedish, Spanish, and Greek. It is a matter of fact that usually the /p/ in Greek and Spanish, absolutely and relatively, shows a much shorter duration than the /p/ in Standard Swedish or Finnish. In order to execute the essential element or elements of the p-gesture, a certain minimum of time is required. However, it is quite evident that the p-occlusion in Finnish and Standard Swedish is held longer than necessary, especially following a stressed short vowel in Standard Swedish or in a long consonant in Finnish, in order to be able to produce a good and complete /p/. Therefore everybody will realize that in certain languages, like for instance Swedish and Finnish, temporal gestures or features (quantity) are superimposed on spectral and tonal features in certain positions or, to put it in terms of the gesture theory, spectral and tonal features are coarticulated with temporal gestures, the latter, though, providing the basic and controlling frame of reference.

From a general linguistic view, it can be assumed that, in principle, temporal and tonal phonological features are autonomous in every language. The kind and degree of mutual relationships, of course, may vary from language to language. For instance, the tonal feature of sentence accent or emphasis does not increase segment duration in Danish (Thorsen 1980), in German it may do it optionally (Bannert 1982b), whereas in Swedish it will do so obligatorily. Danish does not show the decrease of vowel duration as a consequence of the increasing number of unstressed syllables in the stress group (Fischer-Jørgensen 1982) which is well witnessed in many languages.

Another problem with the model of the primacy of Fo (Lyberg 1981) where the duration of the stressed vowel is calculated from the change of Fo over this segment is the fact that, as a consequence, no segment duration can be calculated when there is no change of Fo over a given segment. The Fo-declination throughout the utterance is not considered as an Fo-change in this respect. A non-changing Fo is to be found in cases where Fo-points of the same level are concatenated low or high. Take, for instance, long compounds in Standard Swedish like bòstadsbyggnadsprogramkommitté where the low end of the word accent fall is connected low with the beginning of the rise of the phrase accent in the penultima or sentences with several syllables between accentuated ones: Det var ju Pér som skulle ha skrivit brévet. Here the tonal concatenation is high between the high tone of the phrase accent in Pér and the fall of the word accent in bré-. If, then, vowel duration cannot be calculated in such cases, let alone the duration of consonants, durations, in a model of speech processing, have to be taken from somewhere in order to assign typical and correct durations to all these segments. Even this consideration points to a solution treating durations and tonal movements as autonomous units.

Apart from the autonomous temporal and tonal features that, alone or in coarticulation, lay out their basic patterns in the time and frequency dimensions, one can observe various mutual effects of the prosodic features on each other at the phonetic level. These effects are temporal-to-tonal and tonal-to-temporal as well. A rising tonal movement usually takes more time than a falling one (cf. Ohala and Ewan 1973, Sundberg 1979; Elert 1964). This effect, however, is rather small compared to the total segment duration.

One effect of duration on Fo is due to speech tempo. Increased speech tempo leads to shorter segment durations and therefore there is less time available to execute the tonal gestures in addition to the spectral and temporal ones. Increased speech tempo, in general, increases Fo globally throughout the utterance and the range of Fo-variation is decreased (cf. Gårding 1975).

Outline of a model for prosody

The evaluation of the results of this study and the data and conclusions of other investigations (e.g. Thorsen 1980, Bruce 1981, Bannert 1982b) leads to the reasonable view that the temporal and tonal structures of an utterance, in one sense, are totally independent of each other. In another sense, however, they are connected and interlocked with each other. The temporal structure is considered primary, thus representing the necessary requirement for the execution or coarticulation of tonal features. Therefore, in a generative model of prosody, both dimensions, time and frequencey, have to be treated as autonomous dimensions, although mutual dependencies are to be found. What is important for the design of a prosody model is realizing that there are some separate temporal and tonal phonological features, thus that neither is derived from the other. The essential parts of the tonal contour of an utterance, expressed as tonal points or tonal movements, however, are projected onto the temporal structure which has been processed without any a priori dependence on tonal features.

Every prosody model represents only one part of a comprehensive speech model. The generation of prosody must not be seen as the last step in the derivation of the speech signal. Given the interlocking of the prosodic features with different linguistic components (pragmatic, semantic, syntactic, morphological, and phonological), it becomes quite obvious that the prosodic features have to be processed in a fully integrated way on several levels (cf. van Wijk and Kempen 1985). In accordance with this view of integration, the treatment of temporal and tonal structure isolated from other features and processes is abandoned and a comprehensive and interactive prosody model will be outlined. It is illustrated in Fig. 4.

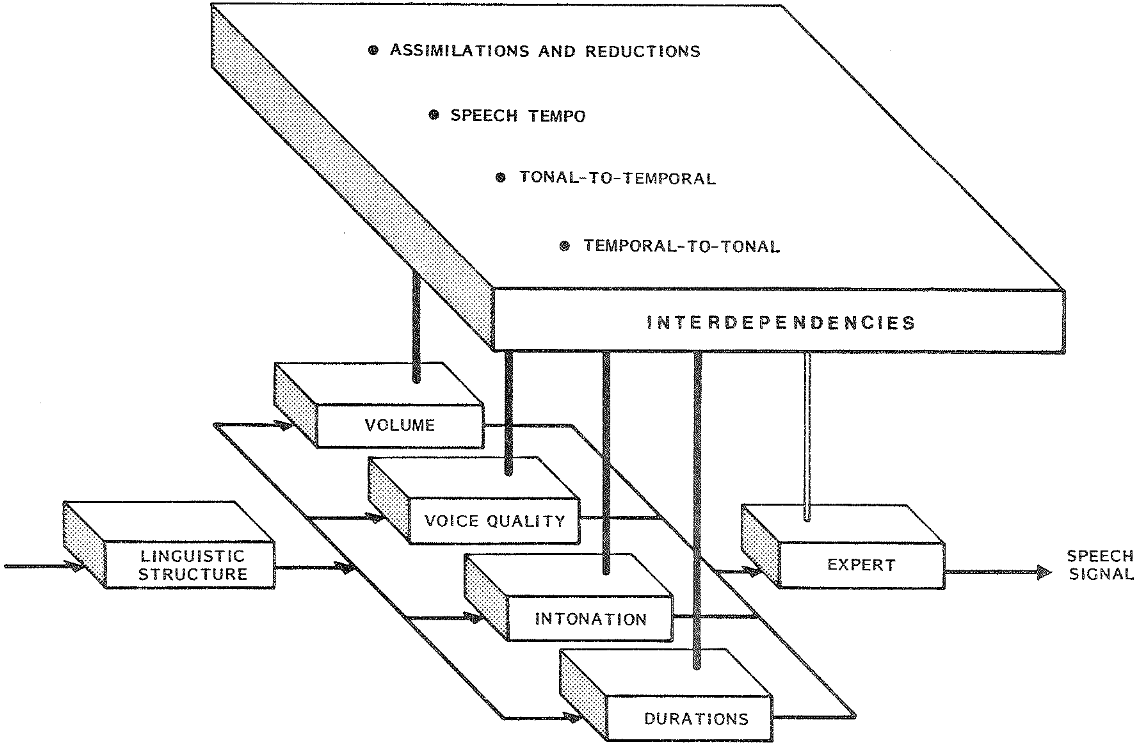

Fig. 4 Outline of a model for prosody where the prosodic structure of utterances is processed in parallel taking into consideration the different factors and relationships affecting the temporal and tonal structures of the output.

As before, the input of the prosody model is a linguistically fully specified string of linguistic units in a phonologically canonical form. The string is completely defined and contains all the necessary phonological (spectral, temporal, and tonal) features including voice quality and volume <5>, as well as the morphological, syntactic, semantic, and pragmatic features. In contrast to some prosody models, it is assumed that all relevant linguistic rules have operated before. Therefore I presuppose that all accent deletions, the assignment of sentence accent, etc. have already been done.

The processing of the information of the input is not done step by step where the output of one step serves as the only input to the next step. In a previous version of a prosody model, the basic temporal and tonal structures of an utterance were processed in a stepwise way, then added and finally, accounting for the mutual effects, modified in the modification component.

The design of the new model is based on the clear distinction between linguistic rules, information, and knowledge of various kinds. All linguistic rules and information including dependencies between features on different linguistic levels which are necessary for generating the prosodic structure of an utterance are available for the processing of the utterance simultaneously and continuously. Obeying the principles of applicability and utility, the rules and information are recalled and used whenever necessary and suitable. Fig. 4 shows the outline of a prosody model designed according to the principle of continuously flowing and complete information processing. Although the basic temporal and tonal structures, and the dimensions of volume and voice quality as well, are generated in separate channels, these sub-processes are effected and controlled all the time by the rhythmical and tonal rules, the information about the mutual effects of time and frequency, about speech tempo, assimilations and reductions (spectral arrangements). As a consequence of this, no further or final modification which otherwise would be necessary after the addition of the basic structure is needed. In the present version of the model, the prosodic structures are generated and processed in accordance with context and all the other relevant factors from the very beginning. It is immediately clear, however, that generating prosody in this on-line model amounts to a very complex process indeed. Nevertheless, the complexity of the processing of the speech signal should not be a deterrent argument against such an approach. On the contrary, it has to be assumed that the present outline of a prosody model is psychologically more realistic than the more simple and linear step-by-step model. In any case, the present design of the prosody model appears to be in rather good accordance with the essential ideas of coarticulation of gestures suggested in the gesture theory of Öhman et al. (1979).

It seems superfluous to remark, of course, that the present outline of a prosody model needs elaborating, completion and testing. However, the model in its present form makes it easier to formulate relevant and interesting questions. It also represents a new test programme in order to, in a coherent and dynamic model, investigate prosodic rules and interrelationships between different features in the time and frequency dimensions. Using speech synthesis by rule, it will be possible to optimize prosodic research by way of direct feed-back. The present model of continuous information processing for the generation of prosody can be connected with the representation of knowledge of artificial intelligence and with expert systems.

FOOTNOTES

<*> This research was supported by the Bank of Sweden Tercentenary Foundation.

<1> These are the high and low points which, in intonation models, are inserted as supporting points by rule with reference to segments or syllables , thus generating the Fo-contour step by step.

<2> For valuable help and discussion I am grateful to Klaus-Jürgen Engelberg, Olle Engstrand, Lennart Nordstrand, Gerhard Rigoll, and Herbert Tropf.

<3> No statistical testing was carried out for several reasons: The number of observations is too small and they could not be collapsed over variables and speakers.

<4> For the purpose of this comparison, it is immaterial that the first segment in position 1 is [an] of kan, whereas the first segment in positions 2 and 3 is [a] of lämna and långa, respectively.

<5> The dimensions of volume (intensity) and voice quality are included and indicated to complete the picture.

REFERENCES

Alstermark M. and Eriksson Y. 1972. Fundamental frequency correlates of the grave word accent in Swedish: The effect of vowel duration. STL-QPSR 2/3, 53-60

Bannert R. 1979. The effect of sentence accent on quantity. Proceedings of the 9th International Congress of Phonetic Sciences, Vol.II, 253-259. Copenhagen

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

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

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

Bannert R. 1984. Towards a model for German prosody. Lund University, Department of General Linguistics and Phonetics, Working Papers 27, 1-36. Also: Folia Linguistica XIX, 321-341

Bannert R. and Bredvad-Jensen A-C. 1975. Temporal organization of Swedish tonal accents: The effect of vowel duration. Lund University, Department of General Linguistics and Phonetics, Working Papers 10, 1-36 and Working Papers 15, 133-138

Bannert R. and Bruce G. 1976. Svenska rytmer i ljud och bild. Umeå University, Department of Phonetics. Publication 10, 2-6

Bolinger D.L. 1958. A Theory of Pitch Accent in English. 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. Lund University, Department of General Linguistics and Phonetics, Working Papers 21, 49-60. Also: Nordic Prosody II, 63-74. Fretheim T. (ed.). Trondheim

Bruce G. 1983. On Rhythmic Alternation. Lund University, Department of Linguistics and Phonetics, Working Papers 25, 35-52

Elert C-C. 1964. Phonological studies of quantity in Swedish. Uppsala

Engstrand O. 1984. Articulatory Coordination in Selected VCV utterances: A Means-End View. Uppsala University, Department of Linguistics. Report 10

Engstrand O., Nordstrand L., and Zetterlund S. 1978. Some observations on the role of prosodic parameters in the perception of phrase structure in Swedish. Phonetics Symposium, 11-13. Nord L. and Karlsson I. (eds). Speech Transmission Laboratory, Royal Institute of Technology, Stockholm

Fischer-Jørgensen E. 1982. Segment Duration in Danish Words in Dependency on Higher Level Phonological Units. University of Copenhagen, Annual Report of the Institute of Phonetics 16, 137-189

Gårding E. 1975. The Influence of Tempo on Rhythmic and Tonal Patterns in Three Swedish Dialects. Lund University, Department of Linguistics and Phonetics. Working Papers 12, 70-83

Gårding E. 1981. Contrastive Prosody: A Model and its Application . Studia Linguistica 35, 146-165

Gårding E., Botinis A., and Touati P. 1982. A Comparative Study of Swedish, Greek, and French Intonation. Lund University, Department of Linguistics and Phonetics. Working Papers 22, 137-152

Gårding E. and Bruce G. 1981. A presentation of the Lund model for Swedish intonation. Nordic Prosody II, 33-39. Fretheim T. (ed.). Trondheim. Also: Lund University, Department of General Linguistics and Phonetics, Working Papers 21, 69-75

Gårding E., Bannert R., Bredvad-Jensen A-C., Bruce G.,and Nauclér K. 1974. Talar skåningarna svenska? Svenskans beskrivning 8, 107-117. C.Platzack (ed.). Lund

Klatt D.H. 1975. Vowel lengthening is syntactically determined in a connected discourse. J. of Phonetics 3, 129-140

Lindblom B. 1979. Final lengthening in speech and music. Nordic Prosody, 85-101. Gårding E., Bruce G., and Bannert R. (eds). Travaux de l'Institut de Linguistique de Lund XIII

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

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

Öhman S., Zetterlund S., Nordstrand L., and Engstrand O. 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

Ohala J. and Ewan W. 1973. Speed of Pitch Change. JASA 53, 345 (A)

Sundberg J. 1979. Maximum Speed of Pitch Changes in Singers and Untrained Subjects. Journal of Phonetics 7, 71-79

Svensson S-G. 1974. Prosody and Grammar in Speech Perception. University of Stockholm, Monographs from the Institute of Linguistics 2

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

van Wijk C. and Kempen G. 1985. From sentence structure to intonation contour. An algorithm for computing pitch contours on the basis of sentence accents and syntactic structure. Germanistische Linguistik 79-80, 155-182. Sprachsynthese, Müller B.S. (Hrsg.). Hildesheim

Zetterlund S., Nordstrand L., and Engstrand O. 1978. An Experiment on the Perceptual Evaluation of Prosodic Parameters for Phrase Structure Decision in Swedish. Nordic Prosody 15-22. Travaux de l'Institut de Linguistique de Lund XIII. Gårding E., Bruce G., and Bannert R. (eds). Lund