Preprints of Papers for the Working Group on Intonation
The XIIIth International Congress of Linguists, Tokyo, 1982

6

SENTENCE INTONATION IN DANISH

Nina Thorsen

Institute of Phonetics
University of Copenhagen

Introduction

This is a summary report of analyses of intonation in two sets of utterances of greatly varying length. For reasons of space, I shall refer extensively to two papers (21, 22) where detailed accounts of the material and procedure can be found. I shall also chauvinistically refer only very briefly to work on other languages.

I. Sentence intonation in short utterances

I have previously presented data and arguments to support a view of the composition of Standard Danish intonation in terms of a layered system of simultaneous, non-categorical components (18, 19). I have also argued that although the stress group and sentence intonation components are highly interactive on the concrete articulatory and physical level (where the stress group pattern is subject to quantitative variation), there may be a level in the speech production and perception processes where the two can be viewed as invariant entities (20). Figure 1 illustrates three stages in the superposition process, as applied to a concrete utterance, and in Thorsen (18, p. 73) an example is given of the inverse process, i.e. the rational decomposition of the fundamental frequency (Fo) course in an utterance into its constituent components. - Janet Pierrehumbert (13) takes a different view of the representation of Danish intonation, in terms of a sequence of (mono- or bi-tonal) pitch accents. In a forthcoming paper (23) I argue that such a representation is hardly descriptively adequate.

Though subject to quantitative variation, the stress group pattern is a qualitatively constant and recurrent entity, i.e. the course of Fo through the unstressed syllables is largely predictable from the stressed ones; a definition of sentence intonation contour as the course described by the stressed syllables of an utterance is therefore appealing. This concept of sentence intonation is different from current topline or baseline concepts: toplines connect local Fo maxima (3, 4, 8, 16) and baselines connect local Fo minima (4, 8, 10, 12). Neither will serve as a useful determinant of the overall contour in Standard Danish because they will be highly dependent upon stress group composition, as can be deduced from figure 1 and 2: the topline is smooth only if every stress group contains a post-tonic syllable; the baseline will be irregular if the post-tonic syllables transgress the intonation contour - which easily happens if the number of post-tonic syllables exceeds three or four and/or if the slope of the post-tonic syllables in the stress group is steep - as it is with some subjects. The (fictional) line connecting the stressed syllables, however, will always and with every subject exhibit a fairly smooth and gradual course, cf. fig-

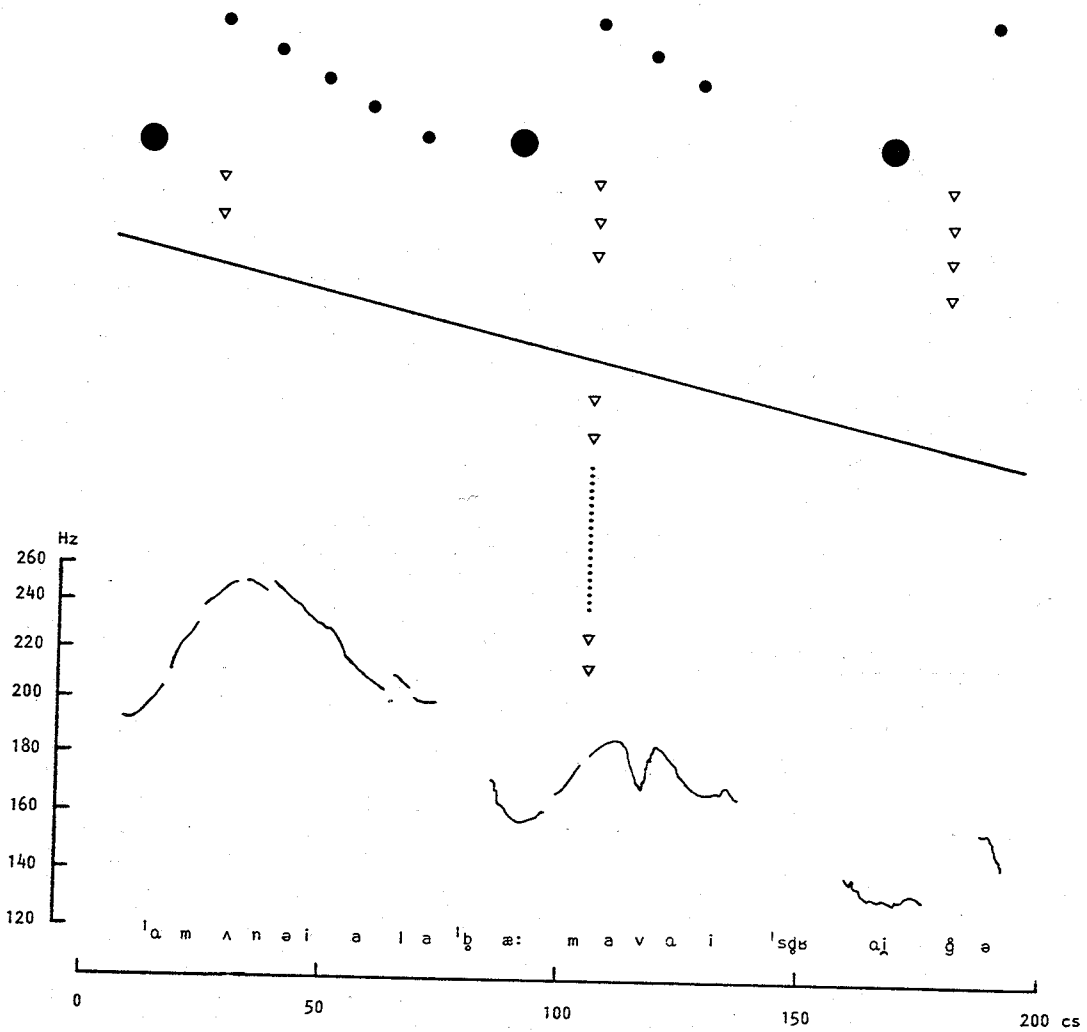

Figure 1

Illustration of the superposition principle in the production of fundamental frequency in an utterance *Ammerne i Alabama var i strejke.* 'The nurses in Alabama were on strike.': Three stress group patterns of different length, superposed on a declining intonation contour will - by way of adjustment rules and microprosodic rules - yield the Fo course shown at the bottom (female speaker).

ure 2. This line is also different from the even less concrete physiologically determined baselines proposed by Fujisaki et al. (7) and Pierrehumbert (13).

The perceptual relevance of the stressed syllables in intonation contour identification was attested in an experiment (19) where subjects had to identify (natural) utterances as interrogative, non-terminal or declarative, solely on the basis of their fundamental frequency course; the segmental structure was identical across the stimuli. It turned out that the distribution of subjects' responses was closely correlated with the stressed syllables - but not, e.g., with any 'terminal contour'.

In short utterances, containing up to four stress groups, the intonation contours approach straight lines whose slopes tend to vary systematically with sentence type and function: terminal declarative sentences have the most steeply falling (unmarked) contours at one extreme, syntactically unmarked interrogative sentences have horizontal contours at the other extreme. In between are found other question types and non-final periods, with a tendency towards a trade-off relationship between syntax and intonation contour: the more syntactic information is contained in a sentence about its interrogative or non-terminal function, the more declarative-like, i.e. the more steeply falling, is its intonation contour and vice versa (17, 19). This tendency was also observed by Alf Bo (1, pp. 82-83) and Otto Jespersen (11, p. 592). A similar trade-off seems to exist also in, for instance, Dutch (5), German (6), English (12), and Swedish (9).

II. Sentence intonation in long terminal declarative utterances

The experiments were originally designed to test a hypothesis that Fo range would be constant over utterances of different length and that accordingly slope would vary inversely with the length of the utterance it spans. The refutation of this hypothesis led to considerations about the interplay between syntax, semantics and intonation.

Two sets of terminal declarative sentences were analysed (referred to in the following as the Thisted-sentences and the Tiflis-sentences, respectively).

1. *Til Thísted.*
2. *Túkke $ skal til Thísted.*
3. *Búster $ skal med bússen $ til Thísted.*
4. *Kísser $ skal med bússen $ til Kílden = i Thísted.*
5. *Lissi $ skal med bússen $ til fésten $ på Kílden = i Thísted.*
6. *Aníta $ skal med bússen $ til fésten = for Kísser $ på Kílden = i Thísted.*
7. *Hútters $ skal med bússen $ til fésten = for Kísser = og Lissi $ på Kílden = i Thísted.*
8. *Knúdsen = og Bítten $ skal med bússen $ til fésten = for Kísser = og Lissi $ på Kílden = i Thísted.*

1. *Til Tíflis.*
2. *Túkke $ skal til Tíflis.*
3. *Búster $ skal med bússen $ til Tíflis.*
4. *Kísser $ skal med bússen $ i nát $ fra Tíflis.*
5. *Lissi $ skal med bússen $ klokken ét = i nát $ fra Tíflis.*
6. *Pýtte $ skal med bússen $ til Thísted $ klokken ét = i nát $ fra Tíflis.*
7. *Hútters $ skal med bússen $ fra kírken = i Thísted $ klokken ét = i nát $ til Tíflis.*
8. *Knúdsen $ skal med bússen $ fra pládsen = ved kírken = i Thísted $ klokken ét = i nát $ til Tíflis.*

The two longest sentences translate 'Knudsen and Bitten are taking a bus to the party for Kisser and Lissi at Kilden in Thisted.' and 'Knudsen is taking a bus from the square by the church in Thisted at one o'clock tonight for Tiflis.' "$" denotes the boundary between major syntactic constituents (NP, VP, complements of time, place, and purpose); "=" denotes minor syntactic boundaries. The stressed vowels are indicated here with acute accents. Note that the syntactic and prosodic boundaries never coincide: the syntactic boundary occurs after the

first post-tonic syllable, except that it occurs directly after *et* and *nat* in the Tiflis-sentences.

I wish to emphasize that the utterances are syntactically unambiguous non-compound sentences. They were uttered in a pragmatically neutral situation (read aloud, out of context), and this is accordingly the only type of speech material I will make any claims about. I should also point out that temporal relations have not been investigated; I am concerned only with the tonal properties of these utterances.

The sentences were recorded six times each by four Standard Danish speakers. The tapes were processed by hardware instrumentation and the Fo curves measured by hand, see further (21, 22). - First of all, standard deviations on mean Fo and time measurements with each subject were small, rarely exceeding 5% of the mean; secondly, average Fo tracings were qualitatively sufficiently alike across subjects to justify the calculation of a grand mean (mean of means) over all four subjects. These grand means are shown in figure 2, where each sentence accordingly represents the average over 24 items (21 only, in the Tiflis-sentences which were recorded three times only by one of the subjects). A description of individual speakers is given in (21) and (22).

In figure 2 the stressed vowels (large dots) are connected with full/broken lines (the intonation contour proper) which are of course fictional in the sense that they are not directly present in the Fo course. The lines are broken in places where I consider prosodic phrase boundaries to be located, see further section 4. below.

1. Range

Range variation and its dependence upon higher and/or lower beginning and end points, respectively, are not the central issue here and a brief summary will suffice: Fo range is NOT constant over utterances of different length; there is an overall tendency for range to increase with increased length but range is not a linear function of utterance length. This is true whether range is defined in terms of the stressed vowels, the first post-tonic vowels, or the absolute Fo maxima and minima. The overall tendency towards increased range is brought about by a combination of higher starting and lower ending points. Stressed vowel end points lower regularly with increased length, until a (physiologically) conditioned floor is reached with utterances of five or more stress groups.

2. Degree of downdrift

Least squares regression line slopes for the stressed vowel data points (the intonation contours) and the first post-tonic vowel data points (the toplines) in sentence 2 through 8 in the two sets are given in table 1. Since the correlation coefficients are generally well above 0.90, such straight line approximations are fairly good representatives of the degree of overall downdrift in the utterances (although they do not capture the finer differences in contour shape, cf. section 3. below): overall slope decreases with increased utterance length, but not linearly so. Furthermore, intonation contour slopes and topline slopes are highly correlated across the utterances within each set ($r = 0.99$ and 0.98 in the Thisted- and Tiflis-sentences, respectively), i.e. steeper intonation contours are accompanied by steeper toplines and less steep contours by less steep toplines. Topline slopes are steeper than the intonation contours, due to the way the low-high interval in the stress group tends to decrease with time, creating a characteristic wedge shape, which is also seen in other languages, for instance Swedish (8). This slope difference is largest in the shorter utterances, however.

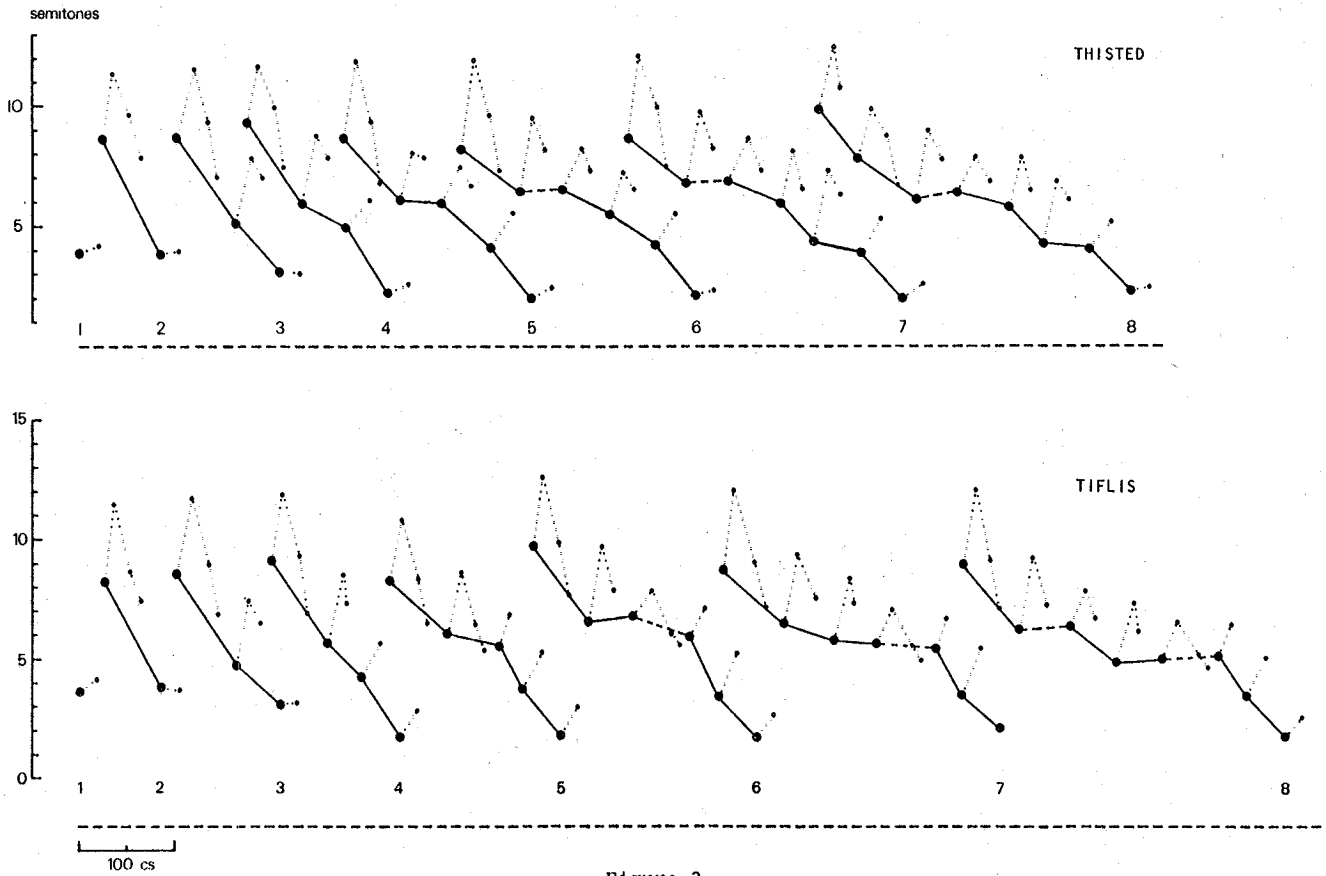

Figure 2

Intonation contours (full/broken lines) and stress group patterns (dotted lines) in two sets of terminal declarative sentences containing from one to eight stress groups. Average over four subjects (mean of means over six recordings). Large dots represent stressed syllables, small dots unstressed ones. In Tiflis-sentence no. 7 the last post-tonic vowel could not be measured with one subject and this point is therefore also missing in the mean presented here. The full/broken line distinction is explained in section II.4. Zero on the logarithmic frequency scale corresponds to 100 Hz.

Table 1

Slopes of least squares regression lines (semitones/sec) and correlation coefficients, calculated on the intonation contours (the stressed vowels) and the toplines (the first post-tonic vowels) in the two sets of sentences depicted in figure 2.

		THISTED	TIFLIS
stressed vowels	2	-7.9	-7.5
	3	-5.2/-1.0	-5.1/-0.99
	4	-4.5/-0.99	-5.4/-1.0
	5	-3.1/-0.98	-3.4/-0.97
	6	-2.2/-0.96	-3.0/-0.94
	7	-2.1/-0.97	-1.9/-0.94
	8	-2.0/-0.96	-1.6/-0.92
post-tonic vowels	2	-11.0	-11.3
	3	-7.6/-0.99	-7.7/-1.0
	4	-5.6/-0.99	-6.4/-1.0
	5	-4.1/-0.98	-4.1/-0.99
	6	-3.5/-0.98	-3.7/-0.98
	7	-2.8/-0.97	-2.3/-0.98
	8	-2.4/-0.96	-2.2/-0.96

3. Shape of downdrift

First of all, I do not think it matters much whether these intonation contours are presented on a linear or (as they appear here) on a logarithmic scale. The 'noise' in the contours is too great to evaluate the adequacy of one scale over the other when - as the case is - their range stays well within the octave. The tendencies outlined below would appear in either format. Secondly, although the average tracings are fairly good qualitative representations of individual subjects' behaviour, some of the variety and complexity exhibited by the four speakers is naturally lost in them.

Several different intonation contour shapes appear: rectilinear declination in the shorter utterances and in parts of the longer ones; asymptotic overall declination in the two longest sentences (though terminated by a sharp fall in Tiflis-8) and in the early part of Tiflis-5 and -7 as well; the second half of Thisted-6 and the middle portions of Thisted-7 and -8 are rather hyperbolic. This variety is NOT an artefact of the averaging (see 21 and 22); nor is the jaggedness of these tracings due to unequal prominence among the stressed syllables: they are all equally prominent perceptually - this was the intention and one which the speakers fulfilled. I would guess that in less planned (spontaneous) speech the "irregularity" would increase rather than decrease.

4. Prosodic phrase boundaries

Figure 2 shows that by a combination of increased range and decreased slope steepness, four stress groups can be accommodated on a smooth and nearly rectilinear intonation contour. At and above five stress groups, the contours are more irregular. The conceptually simplest way to increase the length of the utterance (a) without falling through the floor of the speaker's Fo range and (b) simultaneously preserving a certain declination would be to (partially) reset the contour in appropriate places and maintain rectilinear declinations between resettings, thereby decomposing the sentence intonation contour into a succession of phrase contours, separated by prosodic boundaries. This description may cover some of the utterances, but not all of them, and its inadequacy is even more apparent when one looks at individual subjects. (22) and (23) contain a discussion of the inadequacy of any purely objective criteria to locate the intonation contour discontinuities, as well as a reservation about the perception of prosodic boundaries - as boundaries per se - in intonation contours in syntactically unambiguous non-compound sentences such as these. I was compelled to conclude that we must content ourselves with a description of the contours and, if possible, an explanation for the way they look, because I cannot write the rules that will enable an automatic speech recognition system - which does not "understand" the utterances - to assign the boundaries as I decided to do, according to more or less objective physical and syntactic criteria, supplemented with comparisons of sentences within and across subjects.

The broken lines in figure 2 indicate places where I posit prosodic boundaries: a rise or a sharp fall both constitute an intonation contour discontinuity and initiate a phrase contour. This is intuitively satisfactory with the Thisted-sentences but in Tiflis-6 the two criteria are conflicting. Via comparisons with Tiflis-7 and -8, the boundary is located before the final fall, which means that the preceding phrase contour ends in a rise. To avoid an ultra-short prosodic phrase in Tiflis-8, only the right-most of the two consecutive rises in the contour (between the fourth and fifth, and fifth and sixth stressed syllable, respectively) is recognized as a prosodic boundary (a) because the left one would cut up the place complement internally (between *kirken* and *Thisted*) and (b) because the right-most one directly precedes the sharp final fall.

The stress group patterns are unaffected by the prosodic phrase boundaries: there is no systematic difference between stress group patterns immediately preceding or succeeding a prosodic phrase boundary and stress group patterns in other positions. And although the prosodic boundaries are affiliated with syntactic ones (cf. section 5. below), the two are not coterminous: the prosodic stress group preceding the intonation contour discontinuity, i.e. a phrase final stress group, swallows any pre-tonic material belonging to a succeeding syntactic constituent.

5. Prosodic phrase composition

	THISTED					TIFLIS					
	NP	VP	Complements			NP	VP	Complements			
			purpose	place				place	time	place	
no. 6	1	1	2	2	(2/4)	1	1	1	2	1	(3/3)
no. 7	1	1	3	2	(2/5)	1	1	2	2	1	(4/3)
no. 8	2	1	3	2	(3/5)	1	1	3	2	1	(2/3/3)

The number of stressed syllables in the major syntactic constituents in sentence 6-8 is shown on the preceding page; the number of prosodic stress groups in each prosodic phrase - as depicted by figure 2 - is indicated in parentheses. There were individual differences between subjects as to the number of prosodic phrases they would cut the intonation contours into but the mean tracings in figure 2 do not violate any differences in principle in subjects' behaviour, and the following observations hold true of all four:

NP and VP are never separated. The reason may be one of several: (1) NP and VP are generally not separated prosodically, (2) a separation would create a prosodic phrase with just one stress group in it - which may be unacceptable, (3) the syntactic boundary before the complement of purpose is stronger than the NP+VP boundary, given the semantic content of the syntactic constituents in these particular utterances. - Likewise, the final complement of place is not separated from the preceding time complement in the Tiflis-sentences. Again, there are several reasons to choose from: (1) a separation would create an ultra-short prosodic phrase (the final prosodic stress group), (2) if ultra-short prosodic phrases are not excluded generally, then at least they will be in this position in the utterance where a break might create a final rise, and that is probably incompatible with the demands on sentence intonation in terminal declaratives, or (3) the syntactic boundary between the time and place complements is weak, given the semantic content of the syntactic constituents in these particular utterances.

A support for an assumption that semantics does have a role to play can be found in sentence no. 6 in the two sets: in Thisted-6 the prosodic boundary is located before the first complement, giving a 2/4 distribution of prosodic stress groups. In Tiflis-6 the prosodic boundary is located after the first complement, giving a 3/3 distribution. The different prosodic structuring cannot be due exclusively to syntactic surface structure. I would hypothesize that the time complement (*klokken et i nat*) in the Tiflis-sentences is less intimately related to the preceding part of the utterance, it is more of a unit apart, and the boundary before it is therefore stronger than any other syntactic boundary in these utterances. Tiflis-6 went as follows: *Pytte skal med bussen til Thisted klokken et i nat fra Tiflis.* Now, if *bussen til Thisted* was one syntactic constituent, equivalent to *the Thisted bus*, then that in itself would explain the prosodic phrasing. But, firstly, this supposed syntactic closeness between *bussen* and *Thisted* is NOT accompanied by a stress reduction on the first element which is otherwise characteristic of close-knit syntactic relations in Standard Danish (like *køre bil* 'to drive a car' and many others). Secondly, the pattern is repeated in sentence 7 which gets a 2/5 division in the Thisted-sentence but 4/3 in the Tiflis-sentence, which runs as follows: *Hutters skal med bussen fra kirken i Thisted klokken et i nat til Tiflis.* 'H. is taking a bus from the church in Thisted ..' and *bussen fra kirken i Thisted* cannot conceivably be one syntactic constituent. On the whole, prosodic boundary location is likely to be a combined result of the length of the syntactic constituents and their semantic content - but it will take a much larger material to explicit the way these two factors are balanced against each other.

The difference in the shape of the contours in the two sets may have the same semantic cause. Sentence (5,) 6, 7 and 8 in the Tiflis-set all end in a rather sharp fall, which makes the final phrase contour stand out distinctly from the preceding part of the intonation contour. If the time complement is semantically less close-knit with the preceding part of the utterance, the boundary before it is stronger, and to signal this we get this rather sharp turn in the contour - which in sentence 6 and 8 is preceded by a very slight rise at the end of the previous phrase contour. Note that the boundary before the time complement would have been a very likely place for a pause to occur (not that it ever

did). With this interpretation we must distinguish two kinds of rises within an intonation contour: a phrase final continuation rise and a resetting whose purpose is to keep the contour within the speaker's Fo range. The two rises may not be physically distinguishable: compare Tiflis-6 and Thisted-6 which did not look very different prior to the boundary assignment. I realize, of course, that this account hinges strongly on the assumption that the discontinuities in these contours are not perceived independently, as prosodic boundaries, properly speaking. Rather, if they are perceived at all, they will be identified by the listener (as they presumably were produced by the speakers) in accordance with his syntactic and semantic interpretation of the utterances (see further 22 and 23). I am naturally also well aware that the procedure I opted for may not be optimal (why is only a clean rise recognized as an upward dislocation of the contour? what is a "sharp fall"? and what is not?), but I have not been able to find any other one which would make these utterances, or rather the description of them, LESS complex.

III. Syntactic and prosodic structures

There are fairly strong grounds for claiming that some prosodic categories are distinct entities in the phonology that do not have an isomorphous relation to surface syntactic structure. First, the same syntactic boundary between constituents with the same semantic content may or may not be affiliated with a prosodic boundary. There are differences between subjects (some introduce more prosodic boundaries than others); there is a dependence on the length of the utterance: compare Tiflis-4 with Tiflis-6, -7 and -8 and note that the time complement which induces a boundary in the longer utterances goes unnoticed by the intonation contour in the shorter one (whether or not Tiflis-5 has a prosodic boundary is debatable, really). Secondly, prosodic stress group boundaries are unaffected by any syntactic boundaries at all (in these syntactically unambiguous non-compound sentences) and when in longer utterances a division of the intonation contour into prosodic phrases is necessitated, this phrasing bears no simple relation to surface syntactic structure, due to the role semantics plays in prosodic structuring. Jørgen Rischel (14, 15) argues in a similar fashion that Danish stress is best represented in a hierarchy (a tree structure) which is not necessarily congruent with syntactic structure. To sum up: prosodic boundaries will be affiliated (but not coterminous) with syntactic ones but syntactic boundaries need not leave any trace in the intonational structure of syntactically unambiguous non-compound sentences. Whether syntactic boundaries can be traced in the time structure or in segmental cues is another question, and one that should be addressed.

How and to what extent the present description would be applicable to spontaneous speech, I cannot say. Spontaneous speech is rarely so syntactically well-formed, and prosodic boundaries may be more evident (also when unaccompanied by pauses) in free speech and may take more and different shapes than encountered here.

References

(1) Bo, A., Tonegangen i Dansk Rigsmål. Studier fra Sprog- og Oldtidsforskning, 164. Branner, København (1933).

(2) Bolinger, D., Intonation as a universal. Proc. Ninth Int. Cong. Ling., pp. 833-848 (1962).

(3) Breckenridge, J. and M.Y. Liberman, The declination effect in perception. Manuscript (1977).

(4) Bruce, G., Word prosody and sentence prosody in Swedish. Proc. Ninth Int. Cong. Phon. Sc., vol. 2, pp. 388-394 (1979).
(5) Cohen, A. and J. 't Hart, On the anatomy of intonation. Lingua, vol. 19, pp. 177-192 (1967).
(6) von Essen, O., Hochdeutsche Satzmelodie. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, vol. 9, pp. 75-85 (1956).
(7) Fujisaki, H., K. Hirose and K. Ohta, Acoustic features of the fundamental frequency contours of declarative sentences in Japanese. Ann. Bull. Res. Inst. Log. Phon., vol. 13, pp. 163-173 (1979).
(8) Gårding, E., Sentence intonation in Swedish. Phonetica, vol. 36, pp. 207-215 (1979).
(9) Hadding-Koch, K., Acoustico-phonetic Studies in the Intonation of Southern Swedish. Gleerup, Lund (1961).
(10) 't Hart, J. and R. Collier, Integrating different levels of intonation analysis. J. Phonetics, vol. 3, pp. 235-255 (1975).
(11) Jespersen, O., Fonetik. Det Schubotheske Forlag, København (1897-99).
(12) Maeda, S., A characterization of fundamental frequency contours of American English. Annali della Scuola Normale Superiore di Pisa, Serie II, vol. IX, pp. 381-408 (1979).
(13) Pierrehumbert, J. Breckenridge, The Phonology and Phonetics of English Intonation. M.I.T. Doctoral Dissertation (1980).
(14) Rischel, J., Compound stress in Danish without a cycle. Ann. Rep. Inst. Phon., Univ. Copenhagen, vol. 6, pp. 211-228 (1972).
(15) Rischel, J., On unit accentuation in Danish and the distinction between deep and surface phonology. To appear in Folia Linguistica.
(16) Sorensen, J.M. and W.E. Cooper, Syntactic coding of fundamental frequency in speech production. In Perception and Production of Fluent Speech (ed.: R.A. Cole), Hillsdale, New Jersey, Lawrence Erlbaum Associates, pp. 399-440 (1980).
(17) Thorsen, N., An acoustical analysis of Danish intonation. J. Phonetics, vol. 6, pp. 151-175 (1978).
(18) Thorsen, N., Interpreting raw fundamental frequency tracings of Danish. Phonetica, vol. 36, pp. 57-78 (1979).
(19) Thorsen, N., A study of the perception of sentence intonation - Evidence from from Danish. J. Acoust. Soc. Am., vol. 67, pp. 1014-1030 (1980).
(20) Thorsen, N., Neutral stress, emphatic stress, and sentence intonation in Advanced Standard Copenhagen Danish. Ann. Rep. Inst. Phon., Univ. Copenhagen, vol. 14, pp. 121-205 (1980).
(21) Thorsen, N., Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish. Ann. Rep. Inst. Phon., Univ. Copenhagen, vol. 14 pp. 1-29 (1980).
(22) Thorsen, N., Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish - Supplementary data. Ann. Rep. Inst. Phon., Univ. Copenhagen, vol. 15, pp. 13-47 (1981).
(23) Thorsen, N., Standard Danish sentence intonation - Phonetic data and their representation. To appear in Folia Linguistica.