ARIPUC 18, 1984, p. 185-242

# INTONATION AND TEXT IN STANDARD DANISH

NINA GRØNNUM THORSEN

*Acoustic analysis of recordings by four Standard Danish speakers shows that each declarative sentence in a text is associated with its own declining intonation contour, but together two or three such contours describe an overall falling slope. Individual sentence intonation contours are steeper and demonstrate greater amounts of resetting between them in a succession of declarative terminal sentences than in a corresponding string of coordinate main clauses. In other words, the closer relation between coordinate structures is reflected in a more coherent or less segregated intonational structure. The results are compared with other languages, and the implications for the abstract representation of Danish intonation are discussed.*

## I. INTRODUCTION

Lehiste (1975, p. 195) hypothesizes

*"... that paragraphs possess a suprasegmental structure that indicates the beginning and end of paragraphs and characterize the body of the paragraph. For example, it may be that the intonation contour applied to a sentence produced in isolation (constituting a one-sentence paragraph) will differ from the intonation contour applied to the same sentence in the beginning, middle and end of a paragraph. In other words, it is possible that a paragraph is characterized by an overall intonation structure to which the intonation contours of its constituent sentences are subordinated. ... A corollary hypothesis is that since speakers and listeners share the same code, listeners are capable of deciding whether a sentence has been produced in isolation or as part of a larger structure."*

Both hypotheses are verified in Lehiste's (1975) experiments. Acoustic analysis of one recording of paragraphs containing one, two, and three sentences (of greatly varying length) by one American English speaker, shows that an isolated sentence is longer than the same sentence in a paragraph context. Fundamental frequency (Fø) in the beginning of a sentence is lower in an isolated sentence than in a paragraph initial one, and higher in paragraph initial position than in medial and final positions. - In a listening experiment subjects were to judge what position a sentence - cut out from its paragraph context - had occupied in the paragraph. Their judgment was significantly better than chance, though not always in accordance with the speaker's intention.

Bruce's (1982) point of departure is an informal listening experiment where two sentences, produced in isolation and in two-sentence texts, are spliced together in various ways. Isolated sentences spliced together, and sentences from texts spliced together in reverse order sound odd, and neither combination constitute a perceptually coherent text. Acoustic (Fø) analysis of the original recordings shows that two sentences in a text are characterized by a continuous overall downdrift, though the sentence boundary is marked by a low, sentence final Fø minimum succeeded by a resetting to the first high Fø peak in the second sentence. This peak is lower than the corresponding Fø peak in the first sentence. Likewise, the first Fø peak is lower in a sentence produced in isolation than in the first sentence of two in a paragraph, whereas its last Fø minimum value is higher in isolation than in the second of two sentences in a paragraph. The difference between a single sentence and a combination of two seems to be signalled globally [i.e. over a temporal scope corresponding to the whole text, NT], except that each sentence in a text also has a final, very low Fø value. - These results are confirmed in acoustic analyses of a larger material with a Southern Swedish speaker. Three sentences (each containing two stress groups) are produced in isolation and in the nine possible two- and three-sentence combinations. They are separated by periods, but each begins with "*Å*" ('And'). The very first and very last Fø values (which occur in unstressed syllables) are low and constant across texts of different durations. Apart from that, local Fø maxima and minima are higher in the beginning of the text than at its end, and they are higher in the beginning of a longer text than in the beginning of a shorter text, whereas the final values are constant across texts of different length. Multi-sentence texts were produced with internal pause(s). Before each such pause Fø drops to the same low Fø value as found in absolute initial and final position (the "floor") and resets at the onset of the succeeding sentence to smoothly continue the overall downdrift of local maxima and minima. When sentences in different positions are compared, isolated sentences turn out to resemble text initial sentences more than text final ones. As Bruce notes (1982, p. 284) this corresponds to Lehiste's (1975) observation that listeners confuse isolated sentences with initial sentences more often than with final ones.

Cooper and Sorensen (1981, p. 85ff) analysed four different two-clause sentences (each clause containing four stressed words), among them a coordinate main clause construction. They found that individual clauses had separate declinations associated with them, but together these individual declinations describe an overall downdrift. They note specifically that the resetting from the lower value at the end of the first clause to the higher value of the beginning of the second clause is independent of any accompanying breathing pause. They hypothesize that resetting between clauses in a complex sentence is more likely to occur in the environment of longer clauses, slower rates of speech, and lower semantic relatedness between the two clauses (p. 97).

Uyeno et al. (1979) investigated Fø in declarative complex sentences in Japanese: a construction with an embedded relative clause, a coordinate main clause construction, as well as a sequence of two simple sentences in succession. Their data show that all three types are characterized by an overall declination with a local resetting at the onset of the second clause or sentence, except when the relative clause is utterance initial, in which case the utterance declines smoothly from the initial high rise. The resetting is greatest before a center embedded relative clause. The point relevant to the present investigation is that coordinate main clause constructions and successions of simple terminal sentences did not differ among themselves.

These results (especially those of Lehiste, 1975, and Bruce, 1982) are all in sharp contradistinction to Nakatani (1975) who reports that listeners, presented with stories read normally and stories spliced together from isolated sentences, had difficulty in distinguishing the stories as normal or spliced. He concludes that speech features observed in isolated sentences can be reasonably generalized to sentences in a coherent context.

The present study is to a large extent similar to those of Lehiste (1975) and Bruce (1982) to whom I owe the idea and general outline of my own procedures. Its aim was to establish how and to what extent sentences in a short text are coupled intonationally, and furthermore to see whether this presumed coupling is dependent upon the (syntactic) relation between the sentences in the text. I shall mainly be concerned here with fundamental frequency, though there is evidence from American English and Japanese that other factors such as length, laryngealization and pause duration are also involved in the production and perception of sentence and clause boundaries, cf. Lehiste (1975, 1979, 1980), Lehiste and Wang (1977), Cooper and Sorensen (1981), and Uyeno et al. (1981).

## II. MATERIAL, SUBJECTS, AND PROCEDURES

### A. MATERIAL

The base is the following three declarative sentences (stressed vowels are indicated orthographically with acute accents here):

(1) *Amánda skal afstéd på cámping.*

'Amanda is going away camping.'

(2) *Hendes mór skal på kúrsus i Týskland.*

'Her mother is taking a course in Germany.'

(3) *Hendes fár skal vándre i Lápland.*

'Her father is going to hike in Lapland.'

In isolation as well as combined in texts of two and three, these sentences could all be uttered in answer to a question about what Amanda and her family are going to do during the summer holidays.

The total number of possible texts that can be constructed from these three sentences amounts to fifteen, which doubles to thirty when two different boundary conditions are introduced. This was an unmanageably large number for subjects to record, if the material was to be mixed with filler utterances of a different syntactic and semantic make-up, which was highly desirable. I decided, therefore, to concentrate on (1) above, i.e. the *Amánda*-sentence, and be sure that one occurred in all the relevant positions.

The following combinations of sentences constitute the material:

1

1 3

2 1

1 3 2

2 1 3

3 2 1

The *Amánda*-sentence is the only one to occur in isolation. Note, however, that both sentence (2) (referred to as the *mór*-sentence in the following) and sentence (3) (referred to as the *fár*-sentence in the following) occur in all three positions in three-sentence texts.

Two different types of text result when the sentences in multi-sentence texts occur in sequences of declarative terminal main clauses versus coordinate main clause constructions, joined by *og* ('and'). Examples are: *Amánda skal afstéd på cámping. Hendes fár skal vándre i Lápland. Hendes mór skal på kúrsus i Týskland.* versus *Hendes fár skal vándre i Lápland, og hendes mór skal på kúrsus i Týskland, og Amánda skal afstéd på cámping.*

The five terminal sentence texts and the isolated sentence were randomized three times, and the total of 18 sentences/ texts were mixed with two materials recorded for different analysis purposes, so they were evenly spread over six pages of reading material. The same procedure was applied to the five coordinate clause texts and the isolated sentence. To avoid any direct comparison of the two types of text they were recorded in separate sessions at least one day apart. In this way the isolated sentence was recorded six times by each speaker, but the texts only three times each. This is a compromise between the demand for as comprehensive a material as possible, a sufficiently large number of recordings of each item by each speaker, and the wish to avoid fatigue effects on the part of the subjects. On a fair number of previous occasions (among them recordings of rather long simple sentences, cf. Thorsen 1983a) subjects' repetitions of the same item have always been rather remarkably constant, with standard deviations on calculated Fø mean values generally below 5% of the mean. I therefore decided that the demand for (the standard, minimum) six recordings of each item would have to yield to the possibility of presenting the texts under two different boundary conditions.

My material differs from Lehiste's (1975) texts ("paragraphs" in her terminology) in that the sentences are all equally long (in terms of the number of prosodic stress groups in each) and in the addition of coordinate main clause constructions. It differs from Bruce's (1982) material in the greater length of each sentence (three versus two prosodic stress groups). And again, where Bruce (1982) so to speak combined a terminal with a coordinate clause construction (by separation of the sentences with periods but commencing each sentence with 'and'), these two conditions are differentiated in the present material.

There is no conflict in Lehiste's (1975) choice of term ("paragraph") and mine (and Bruce's, 1982) ("text"). The present sentence combinations are both paragraphs and texts; at least if one does not claim that the total six pages of reading material constitutes a text. In that case my "texts" would only have the status of paragraphs. However, the six pages could lay no claim on any semantic or pragmatic coherence, and hardly deserve the designation as a text properly speaking.

## B. SUBJECTS

Four phoneticians, two males (NRP and JBC) and two females (GB and NT (the author)) read the material, in two sessions on separate days, as mentioned above. They all speak a form of Standard Copenhagen Danish. Their style of speech during the recordings can be characterized as fluent and conversational but distinct.

## C. TECHNICAL PROCEDURES

The recordings were made with semi-professional equipment (Revox A-77 tape recorder, Sennheiser MD21 microphone) in a quasi-damped room at the Institute of Phonetics on Agfa PE39 tape, at 7½ i.p.s.

The tapes were processed by hardware intensity and fundamental frequency meters (F-J Electronics) and registered on a mingograph (Elema 800) at a paper speed of 100 mm/s. By adjustment of the Fø meter zero-line to the lower limit of the subject's voice range and full exploitation of the record space (about 80 mm) of the mingograph galvanometer, a measuring accuracy of 1 Hz for the males and 2 Hz for the female speakers is attained.

In unidirectional Fø movements in the stressed vowels, the beginning and end point were measured, according to a procedure outlined in Thorsen (1979, p. 63-66). In bidirectional movements, the turning point was measured as well. The unstressed vowels in the sentences were represented by a single Fø point, the midpoint (in time as well as frequency), which was an uncontroversial procedure since most of these vowels have monotonically falling Fø movements. (The first post-tonic syllable may be rising-falling, in which case Fø is measured at the peak.) The Fø movements are so short and slight in the unstressed syllables that it does not much matter which point you choose as the one measuring point: the initial, medial, or final (or any other) value. Changing the location of the measuring point will transpose all the unstressed vowels by very nearly the same (negligible) amount upwards or downwards, relative to the stressed vowels in the figures presented below. There is one instance where two unstressed vowels cannot be segmented. *skal af(stéd)* is pronounced without any intervocalic consonant [sg̊a a], and the measuring points are assigned time coordinates one quarter of the distance from the onset and offset of voicing in the long vowel sound. The distance in time of each measuring point from the first Fø value in the sentence or text was likewise measured. In the terminal sentence texts, time was set to zero at the first measuring point in each sentence, and the pause between sentences (defined as the time interval where no acoustic energy is registered) measured separately. In the coordinate main clause texts, time was measured cumulatively from the onset of the first sentence. Fø and time measurements were averaged over the three recordings by each subject (six recordings in the case of the isolated sentence). Average Fø values were converted to semitones (re 100 Hz) and average tracings drawn. No correction was attempted for intrinsic Fø level differences between the high stressed vowels of the *mór*-sentences and the low stressed vowels of the *fár*- and *Amánda*-sentences.

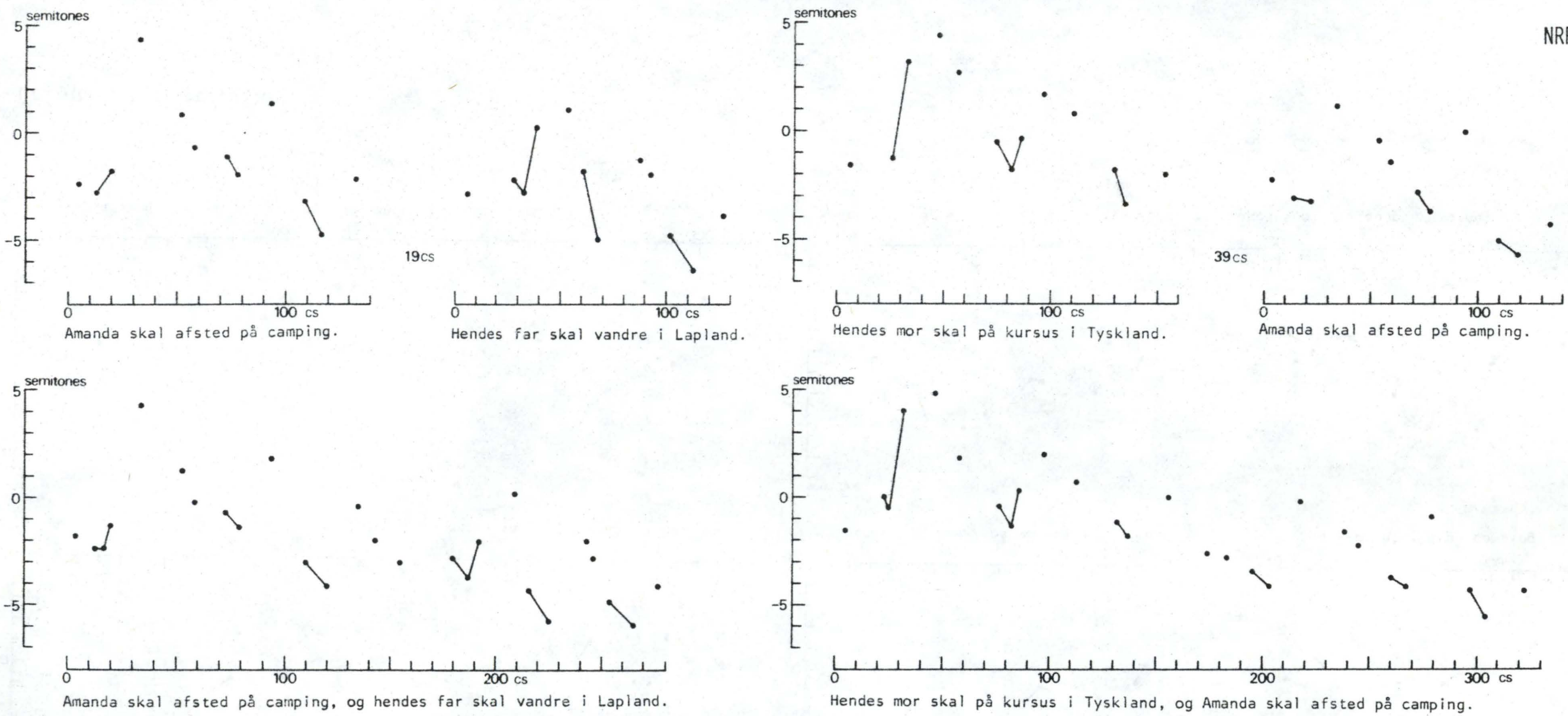
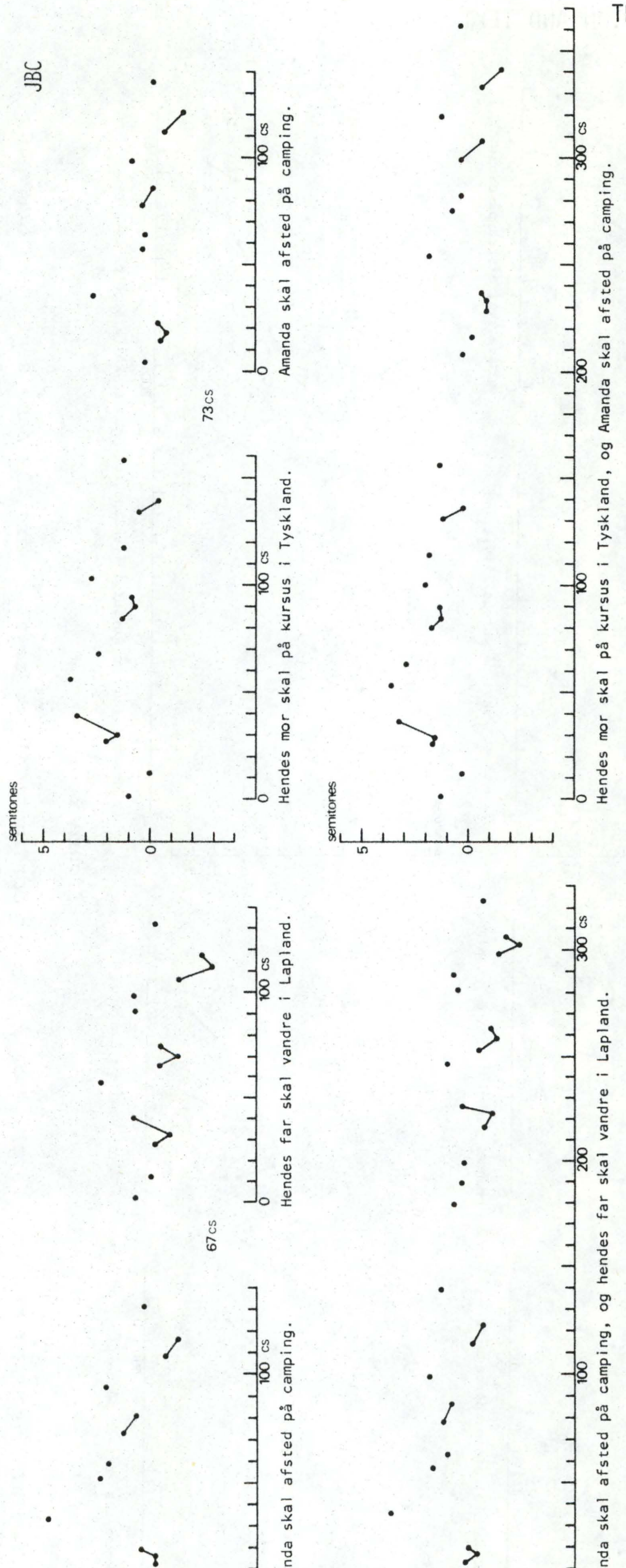
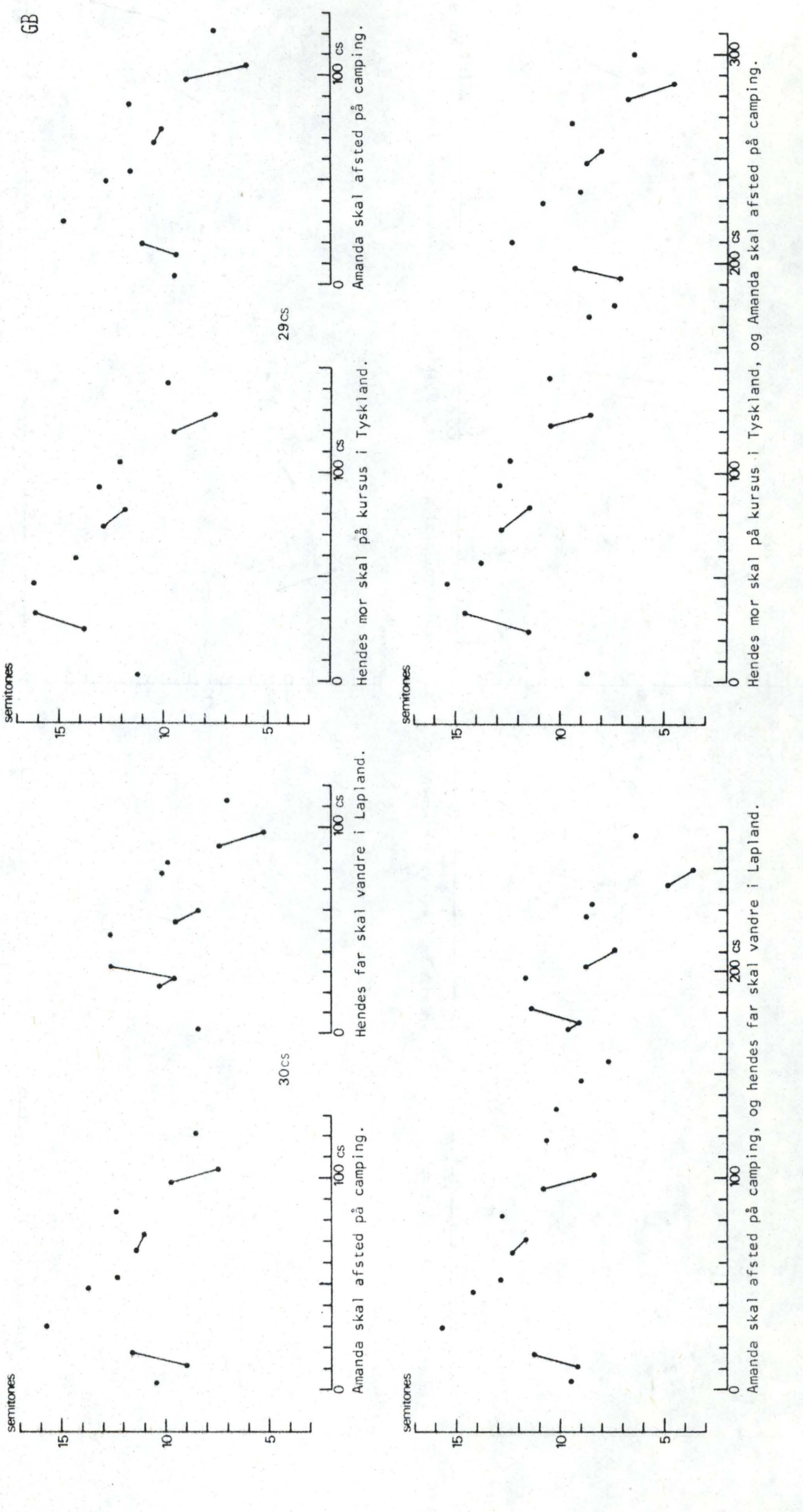



*Figure 1*

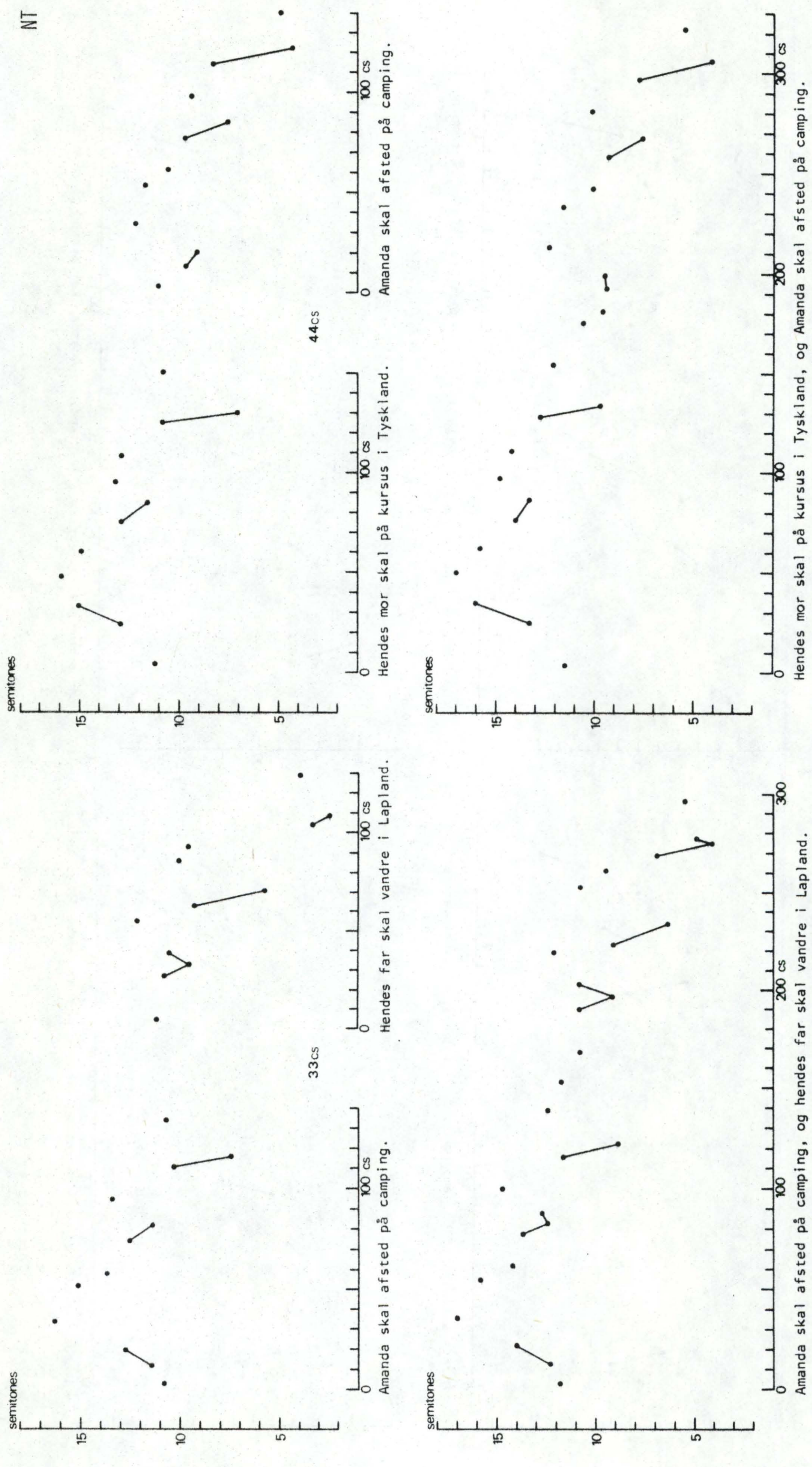
*Fundamental frequency tracings (average over three recordings) of two texts, where the same sentence/clause occurs in text initial position (left side) and text final position (right hand side). Sequences of declarative terminal sentence above (with indication of the duration of the pause between sentences) and coordinate clause constructions below. Stressed vowels are indicated with full lines in the graphs, unstressed vowels are represented by points. Zero on the logarithmic frequency scale corresponds to 100 Hz. Four speakers, two males (NRP and JBC) and two females (GB and NT) - identified at the top right of each set of graphs.*




*(figure 1 continued)*




(figure 1 continued)




*(figure 1 continued)*

# III. RESULTS

## A. PRESENTATION OF THE DATA

In figure 1 is shown average tracings of two pairs of two-sentence texts, where the *Amánda*-sentence occurs in initial and final position, respectively. This is to illustrate the relation between the averaged "raw" data and the data reduced simplified representations in figure 2 (upon which the account of the results and the conclusions are based) and to give a more complete picture of similarities across (and differences between) individual speakers. The stressed vowels are indicated in figure 1 by lines connecting two or three points.

Although the main trends in the data are apparent in figure 1 (and of course also in the tracings of the averaged "raw" three-sentence texts, which are not shown here), a simplification is called for for a number of reasons. The three different sentences are not immediately comparable, due to the distribution of stressed vowel qualities, i.e. rather high vowels in the *mór*-sentence and rather low vowels in the *Amánda*- and *fár*-sentences, which makes the intonation contour too high in the *mór*-sentence relative to the other two, ceteris paribus. Different initial consonants in the stressed syllables within each sentence also obscure the picture. The [m] in *mór* and *Amánda* lowers the initial Fø value compared to the vowels in the following two stressed syllables which begin with unvoiced obstruents. Contrarily, the [f] in *fár* raises the initial Fø value considerably compared to the other two stressed vowels in that sentence. (See further Jeel, 1975, and the references there.) Both these difficulties could of course have been avoided in the construction of the material, but only - I feared - at the expense of the naturalness of the sentences and texts (to the extent that they can lay any claim to naturalness at all). As it is, these "extraneous" factors are alleviated if I cut and splice the texts (in the illustrations) so that initial, medial and final *Amánda*-sentences are presented together in pseudo-texts, and likewise for the *mór*- and *fár*-sentences. In this way we get a clearer impression of the relation between sentence intonation contours (a) in different positions in a text, (b) in texts of different length, and (c) in texts of different types, and it is the relations rather than the absolute slope shapes and values which are interesting in this connection. This last contention also justifies the rather drastic reduction of each sentence into two straight lines, an upper and a lower one. The upper line in figure 2 is the connection between the first and last post-tonic syllable in each sentence. The lower line is the connection of the Fø minimum in the first stressed vowel (the initial value in purely rising movements, the turning point in falling-rising movements) and the final value in the last stressed vowel. This lower line is a reflexion of, and related to, the intonation contour as I have chosen to define it, namely as the line or figure described by the stressed syllables of an utterance. It will be too steep in the *fár*-sentence, and not steep enough in the *Amánda*- and *mór*-sentences, due to influence from the initial consonant in the first stressed syllable. - When I say about vowels that their initial Fø value is "too high" or "too low" and therefore intonation contours will seem "too steep" or "not steep enough", I mean that their physical manifestation does not, presumably, correspond to the way they are perceived. We know that differences in duration and Fø level between stressed vowels of different tongue height, ceteris paribus, are overheard or compensated for by the listener, see e.g. Reinholt Petersen (1974) and Hombert (1977). This compensation is explained with reference to the listener's "knowledge" of those constraints inherent in the speech production apparatus which are responsible for physical differences in sounds which are intended by the speaker to be the "same". From this knowledge presumably derives an expectation to find certain acoustic differences coupled with certain different spectral events, and these differences therefore carry no independent significance and they are disregarded by the listener in the interpretation of what (s)he hears. If the reference to the listener's double function as both speaker and listener (which in a more general perspective is what the motor theory of speech perception is founded on, cf. Liberman et al., 1967) explains the (non) perception of Fø and duration differences between stressed vowels of different tongue height, I think it reasonable to assume that it will also make the listener disregard differences in vowel onset Fø which are caused by the different nature of the preceding consonant. It is certainly my own impression from the recordings that the three different sentences do indeed sound "the same" as far as steepness and shape of the intonation contours are concerned when everything else is equal. This is of course no evidence, and formal experiments should be undertaken to verify or falsify my assumption, but note again that the conclusions to be drawn from the data do not hinge upon its verification since I am more concerned here with relations between sentence intonation contours than with their absolute gestalt.

In figure 2 the simplified data are presented which describe the intonational structure of three pairs of pseudo texts (as defined above), for each of the four speakers and their mean (mean of means). Sentences are identified by their original position; e.g. "1-1" is the first sentence of one, "2-2" is the second sentence of two, etc. The isolated sentence ("1-1") and the two-sentence text ("1-2+2-2") are both presented together with the three-sentence text (the set is complete only in the case of *Amánda*). For ease of comparison "2-2" is shown twice: directly after "1-2" together with the medial sentence ("2-3") and also together with the final one ("3-3").

One type of information is lost in the process of piecing together the figures from parts deriving from different actual texts, namely pause duration. All speakers paused between the terminal declarative sentences, cf. figure 1. Pauses range between 20 and 80 cs. One speaker (JBC) also paused between the coordinate main clauses, but his pauses were only about half as long there (20-50 cs) as between the terminals (50-80 cs).

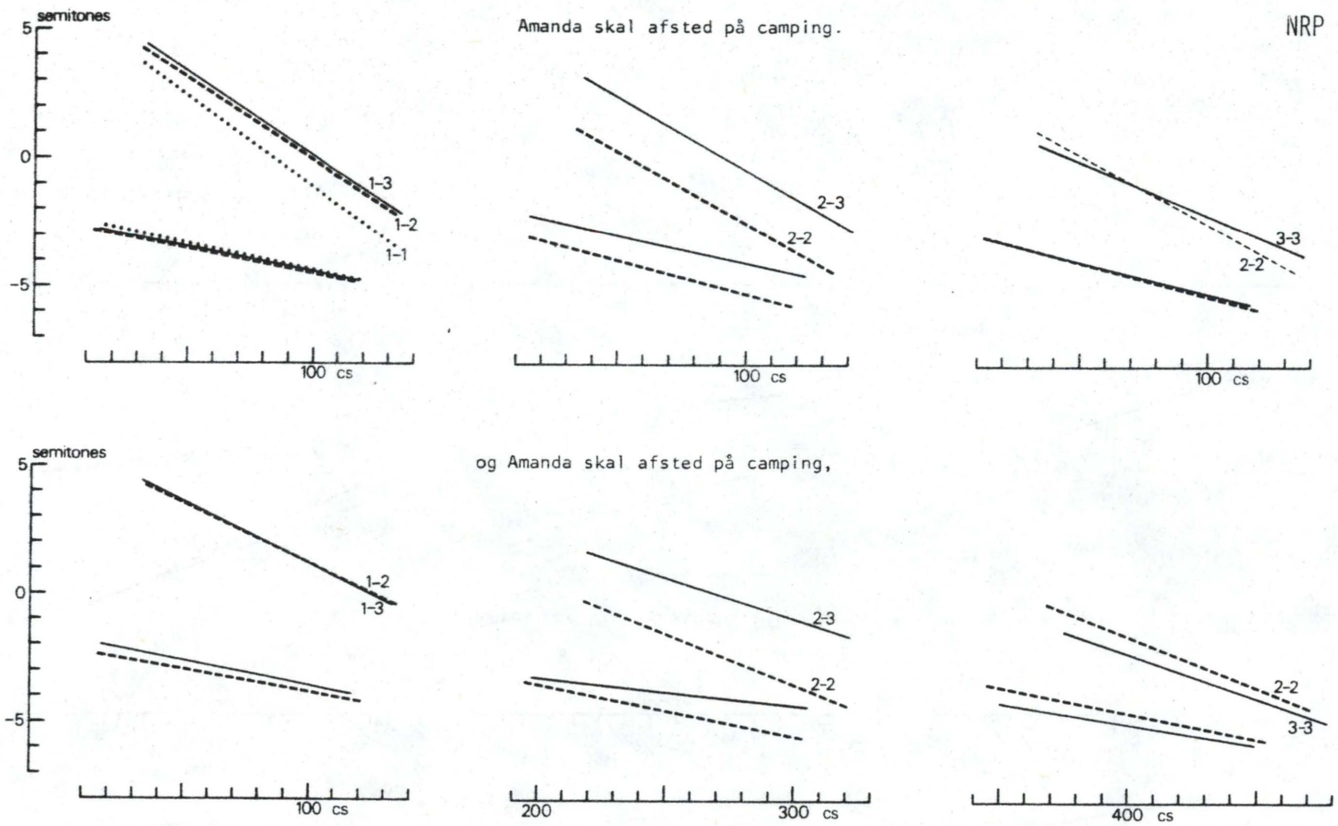
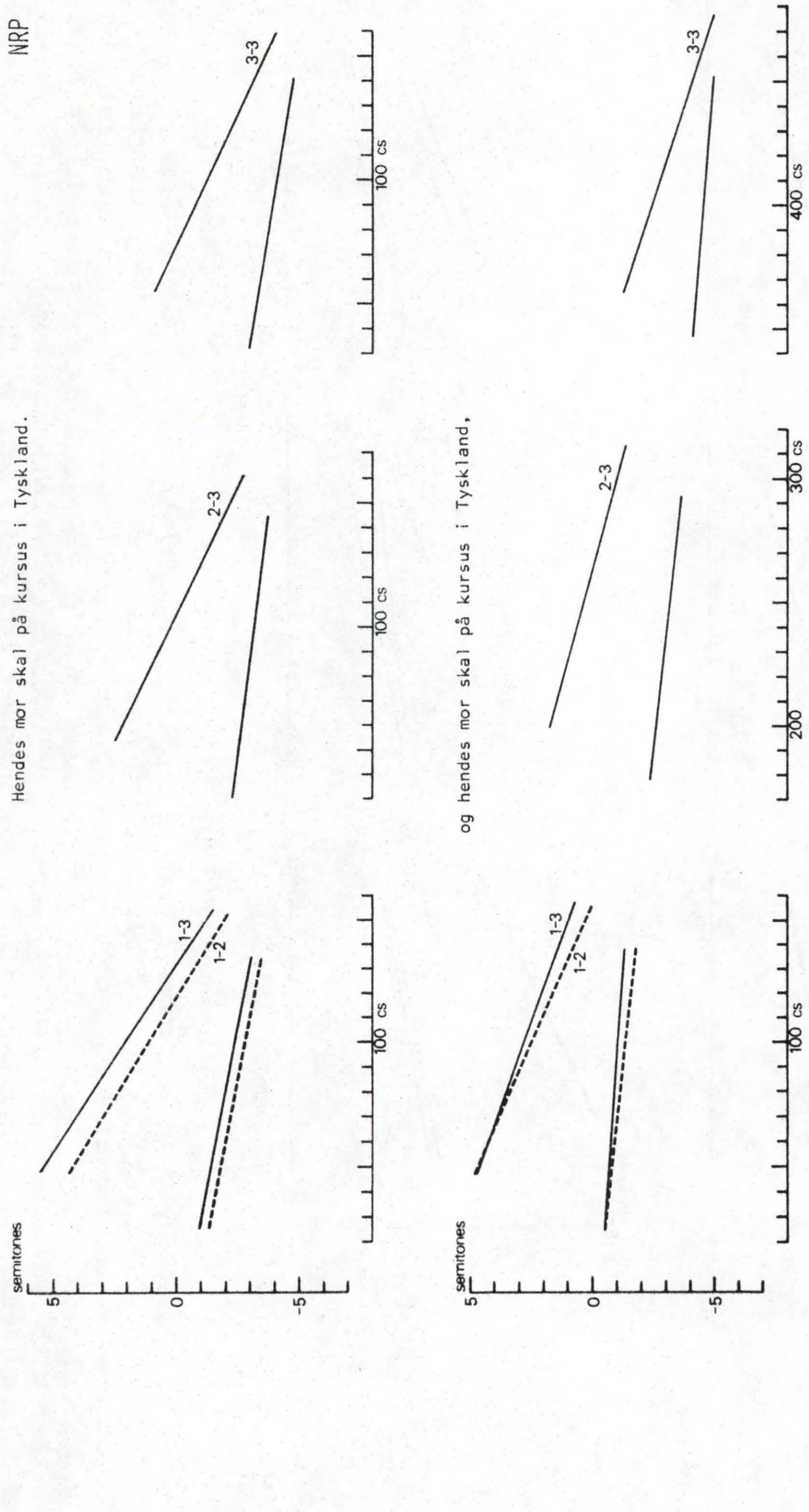
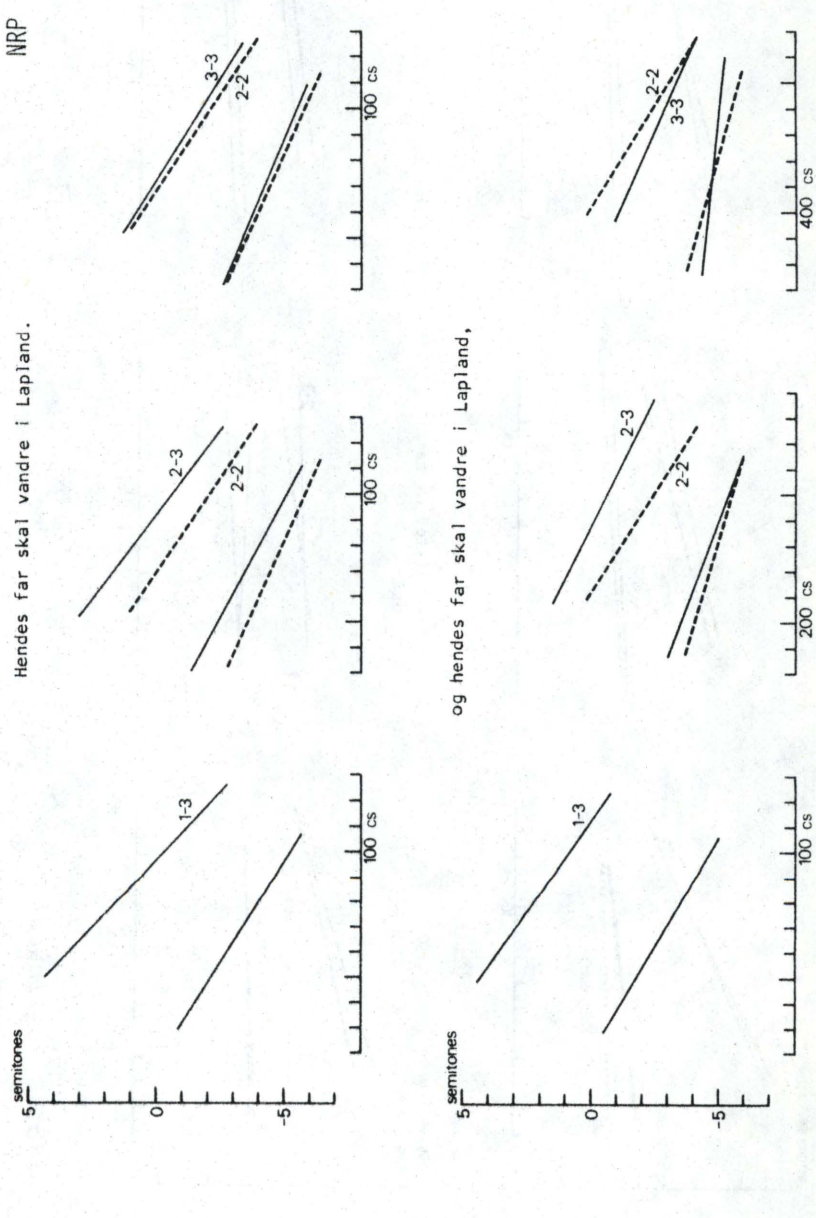



*Figure 2*

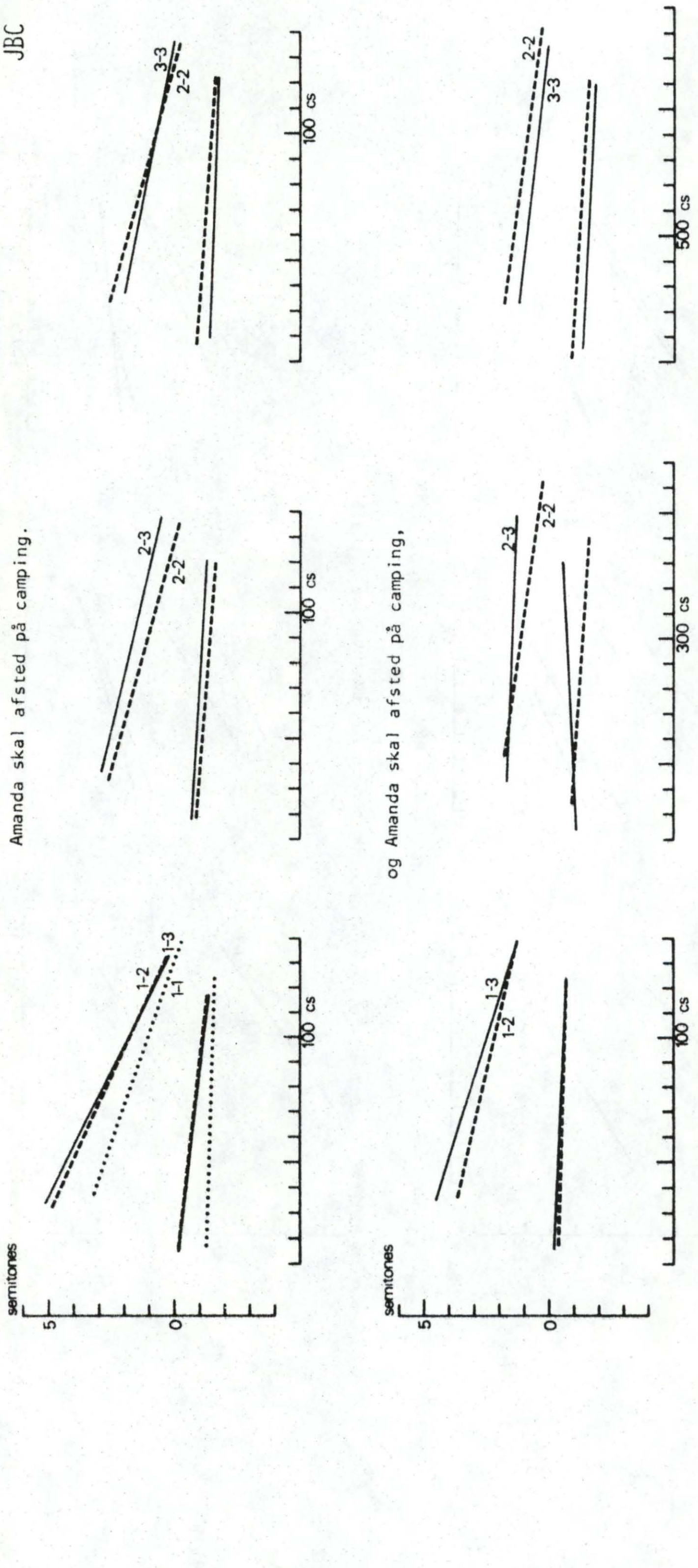
*Upper and lower lines in sentences extracted from text initial (1-2, 1-3), text medial (2-3) and text final position (2-2, 3-3). For an account of upper and lower lines, see the text. There are three different sentences, identified at the top of each set of contours, arranged with sequences of declarative, terminal sentences above and coordinate clauses below. The Amánda .... set is complete, with an isolated sentence (1-1, dotted line), a text with two components (1-2+2-2, dashed line), and a text with three components (1-3+2-3+3-3, full line). 2-2 is depicted twice: together with 2-3 and together with 3-3. Zero on the logarithmic frequency scale corresponds to 100 Hz. There are four speakers, two males (NRP and JBC) and two females (GB and NT) and a grand mean (mean of means) - identified at the top right of each set of graphs.*



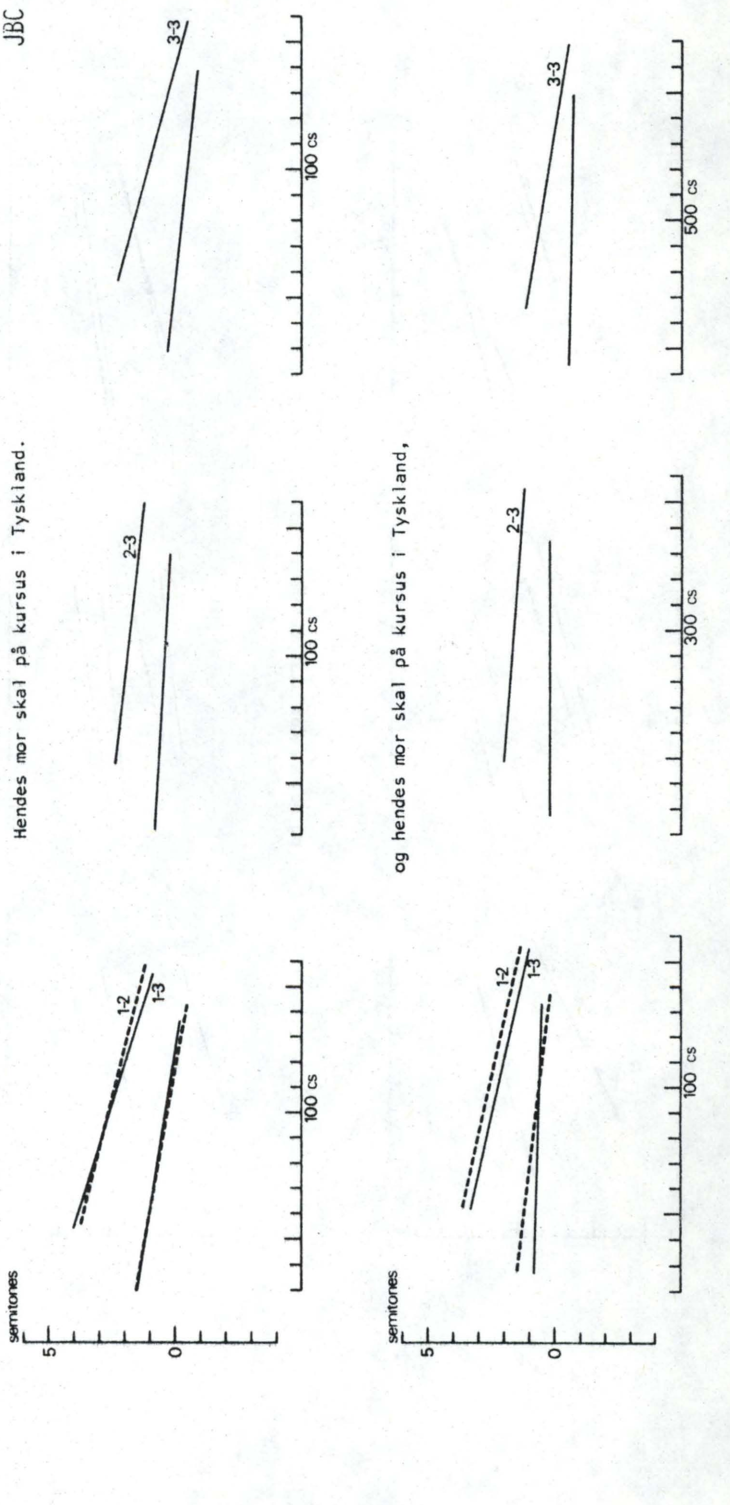

*(fig. 2 continued)*



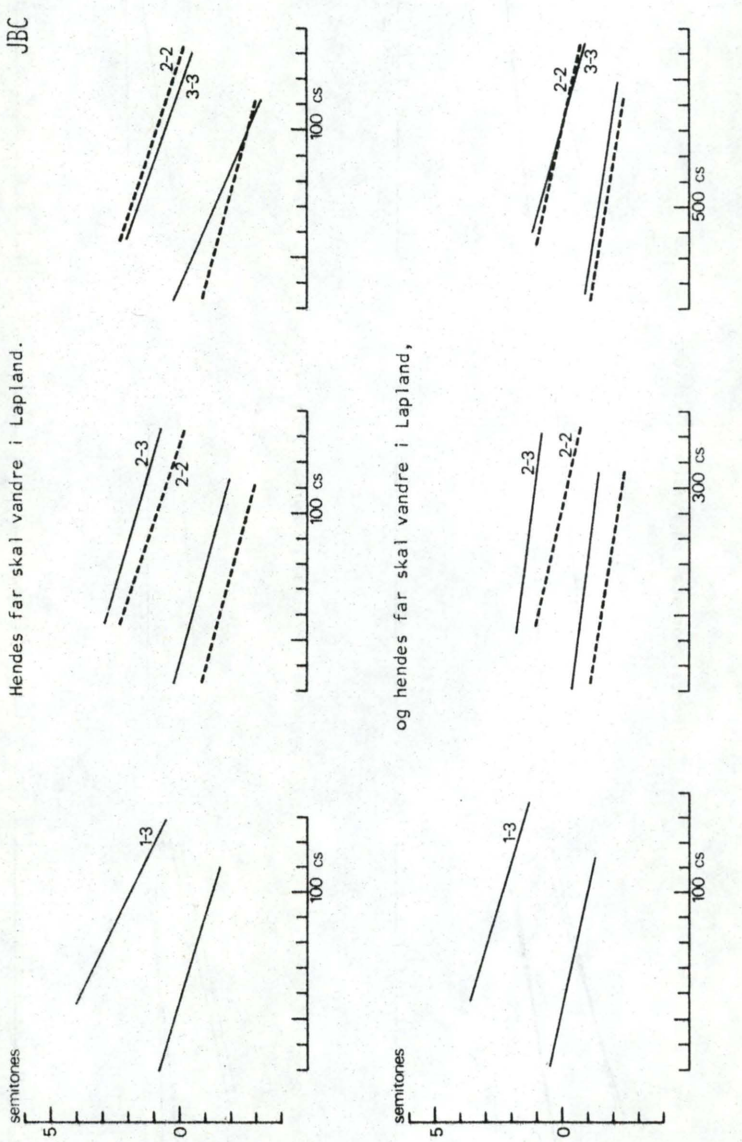

*(fig. 2 continued)*



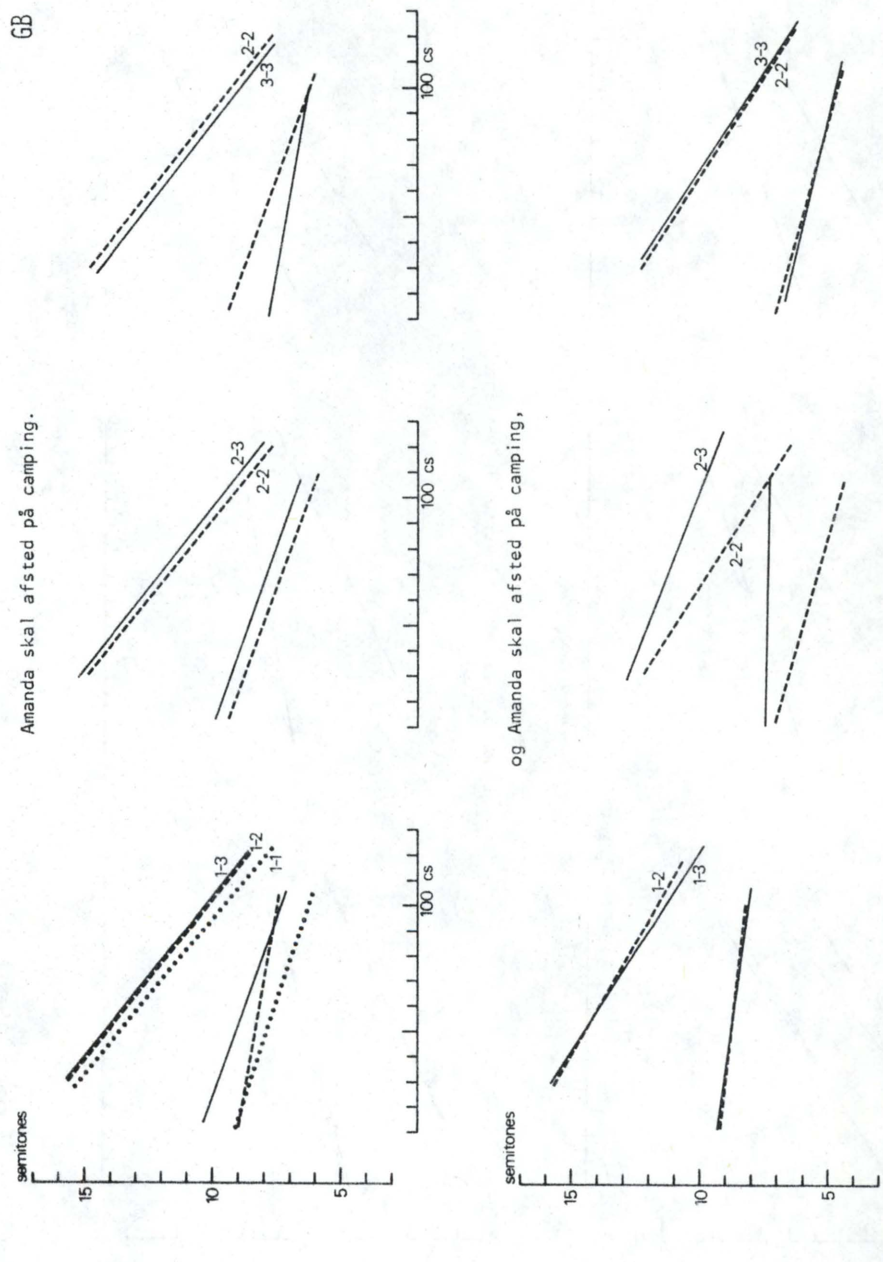

*(fig. 2 continued)*



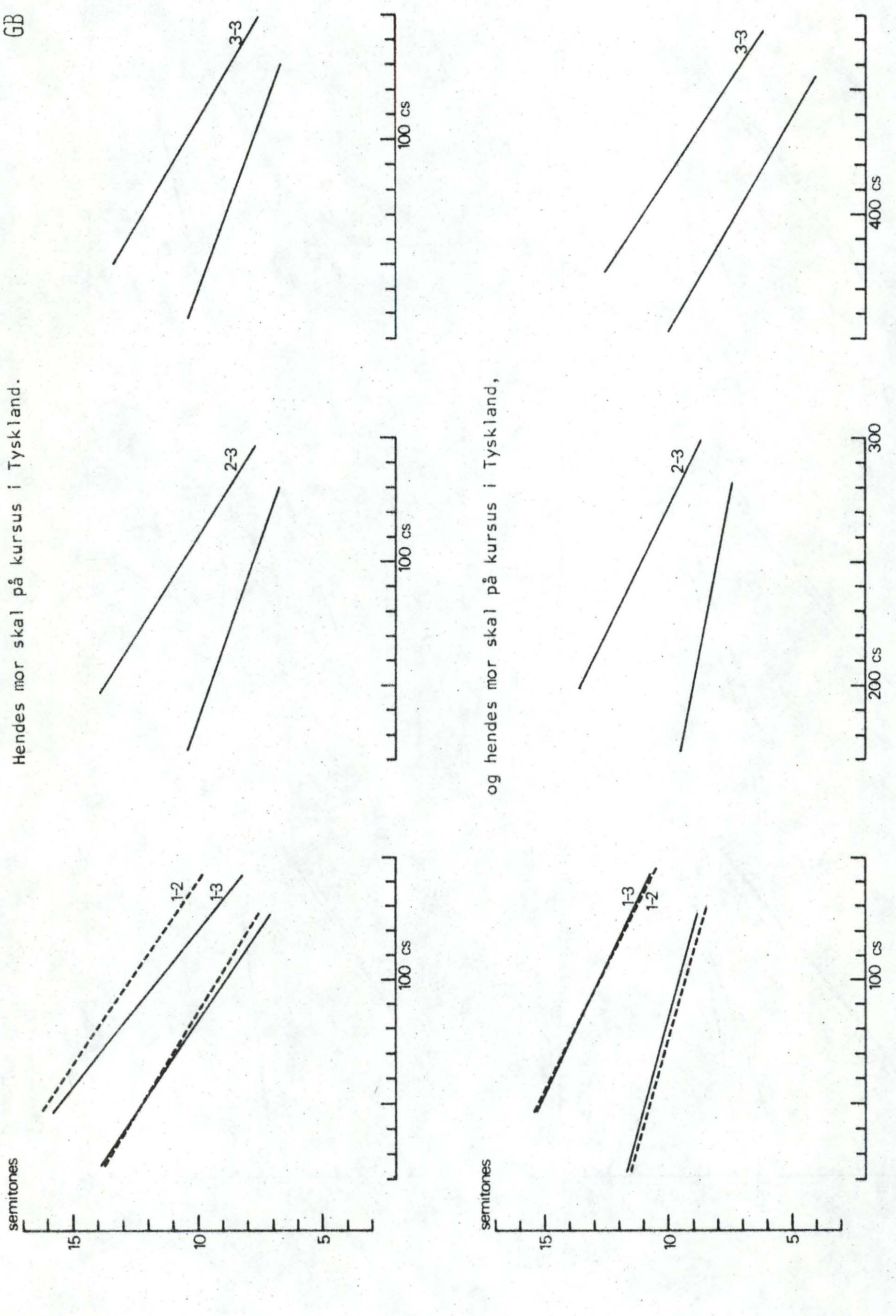

*(fig. 2 continued)*



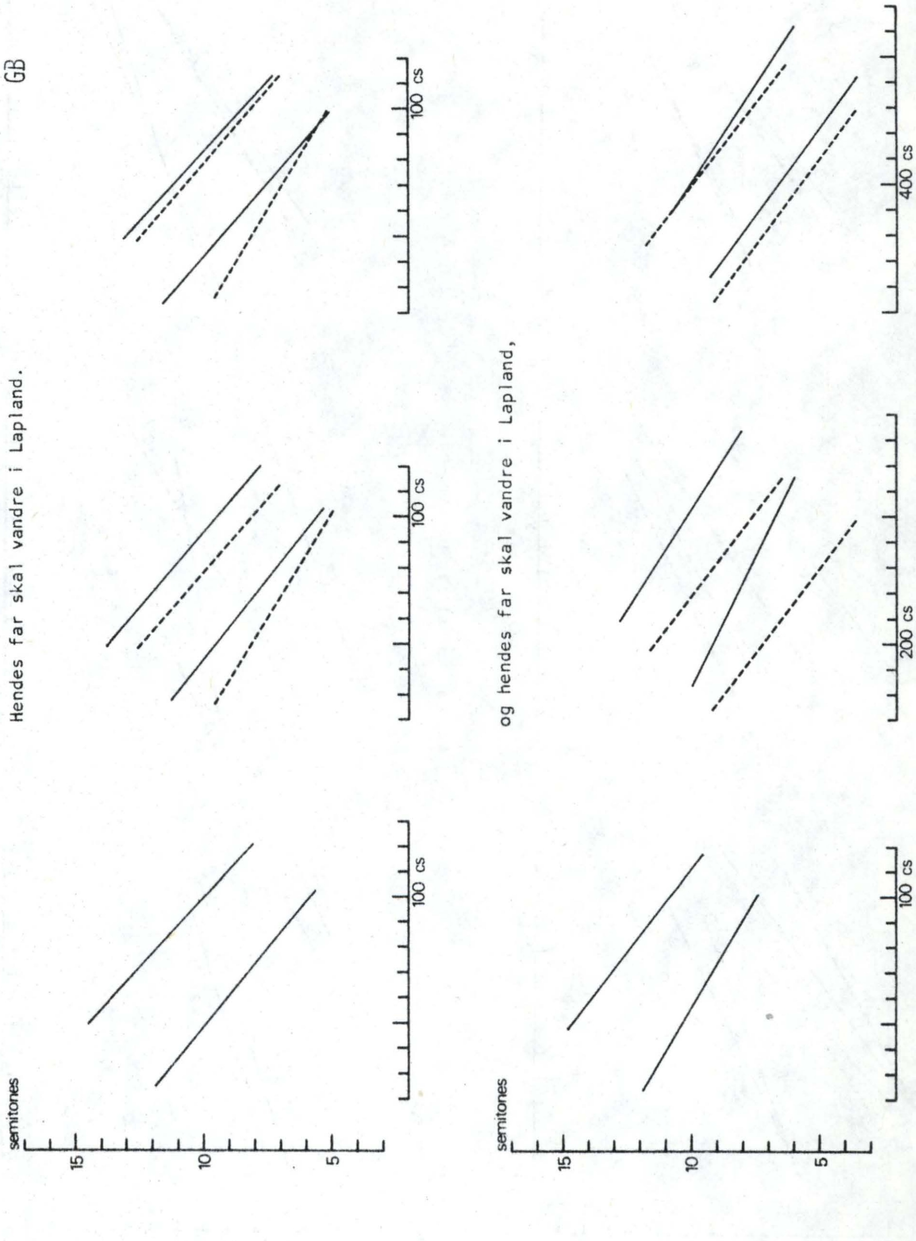

*(fig. 2 continued)*



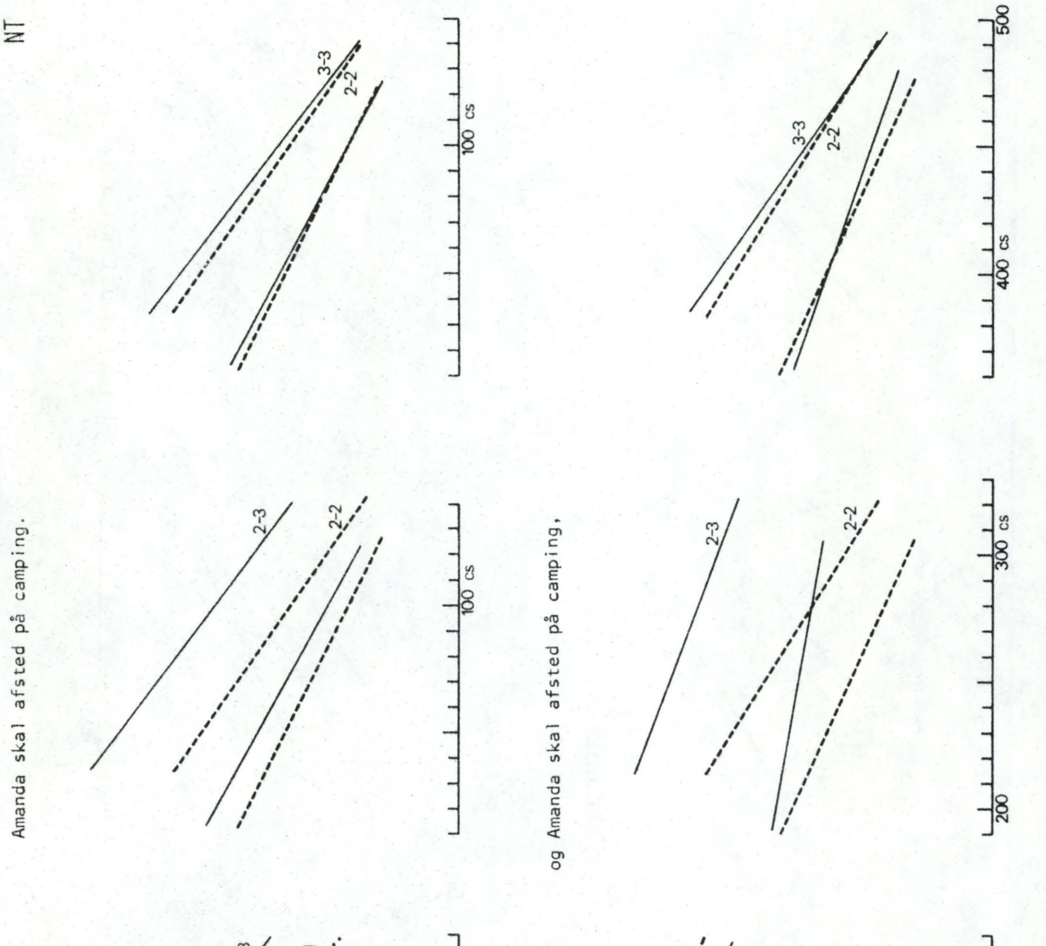

*(fig. 2 continued)*



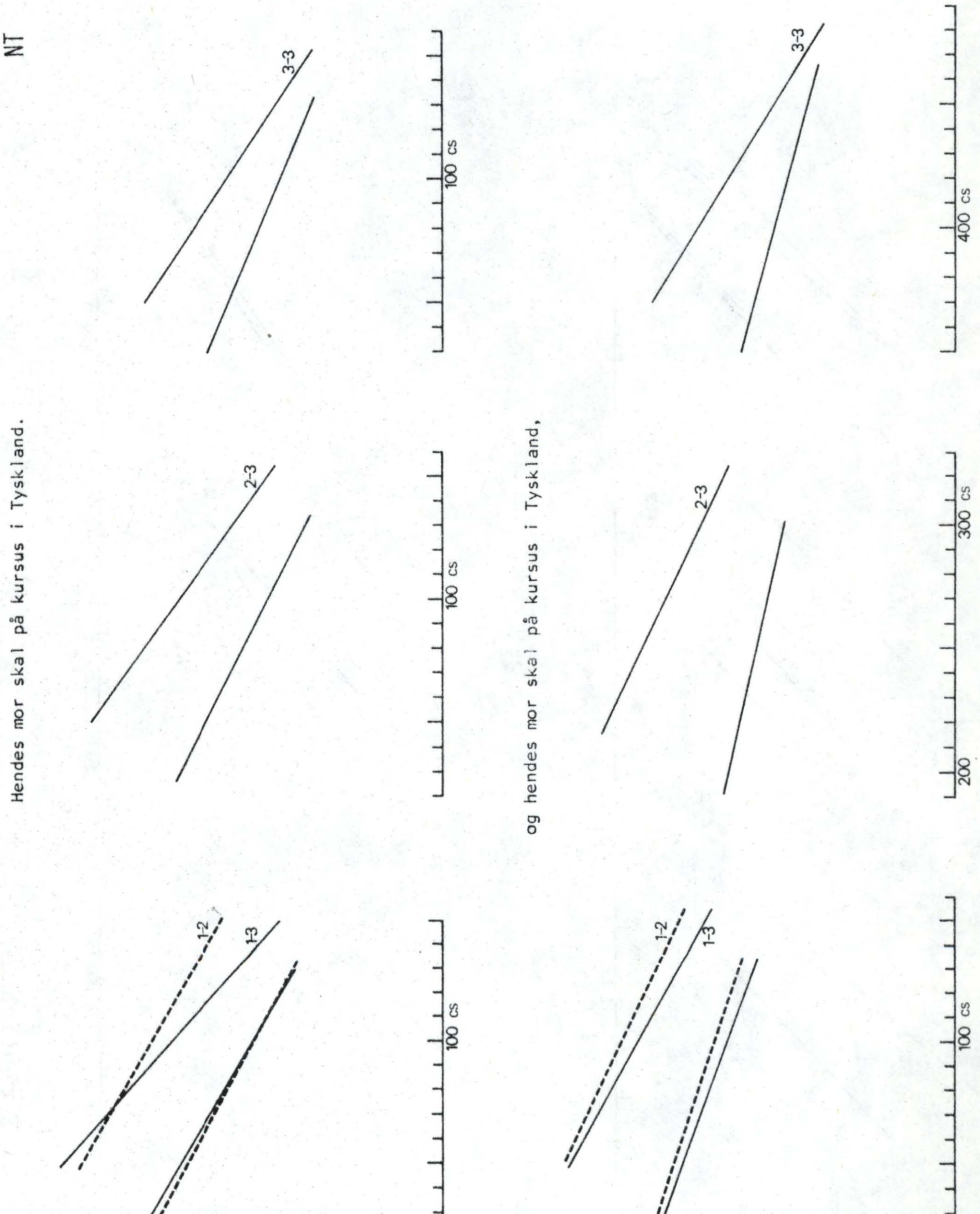

*(fig. 2 continued)*



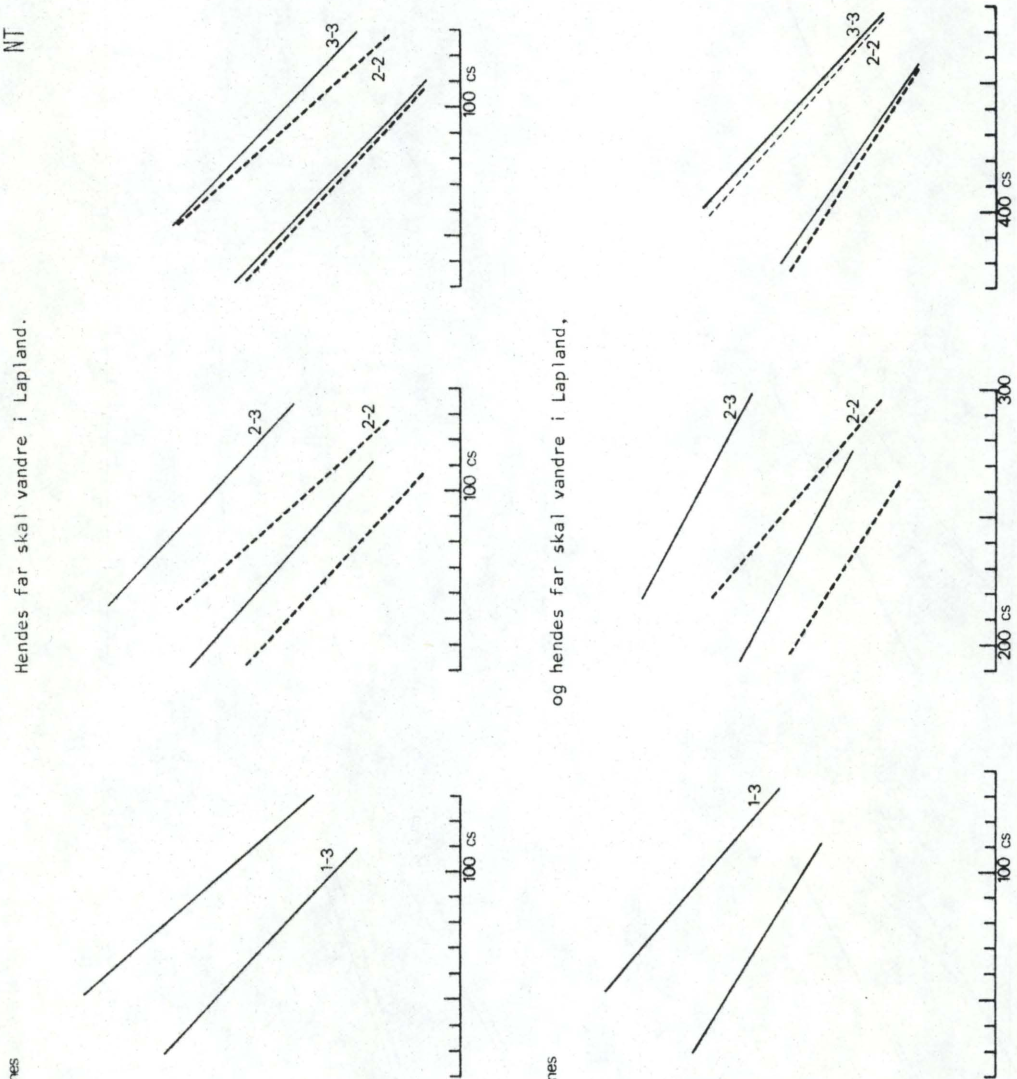

(*fig. 2 continued*)



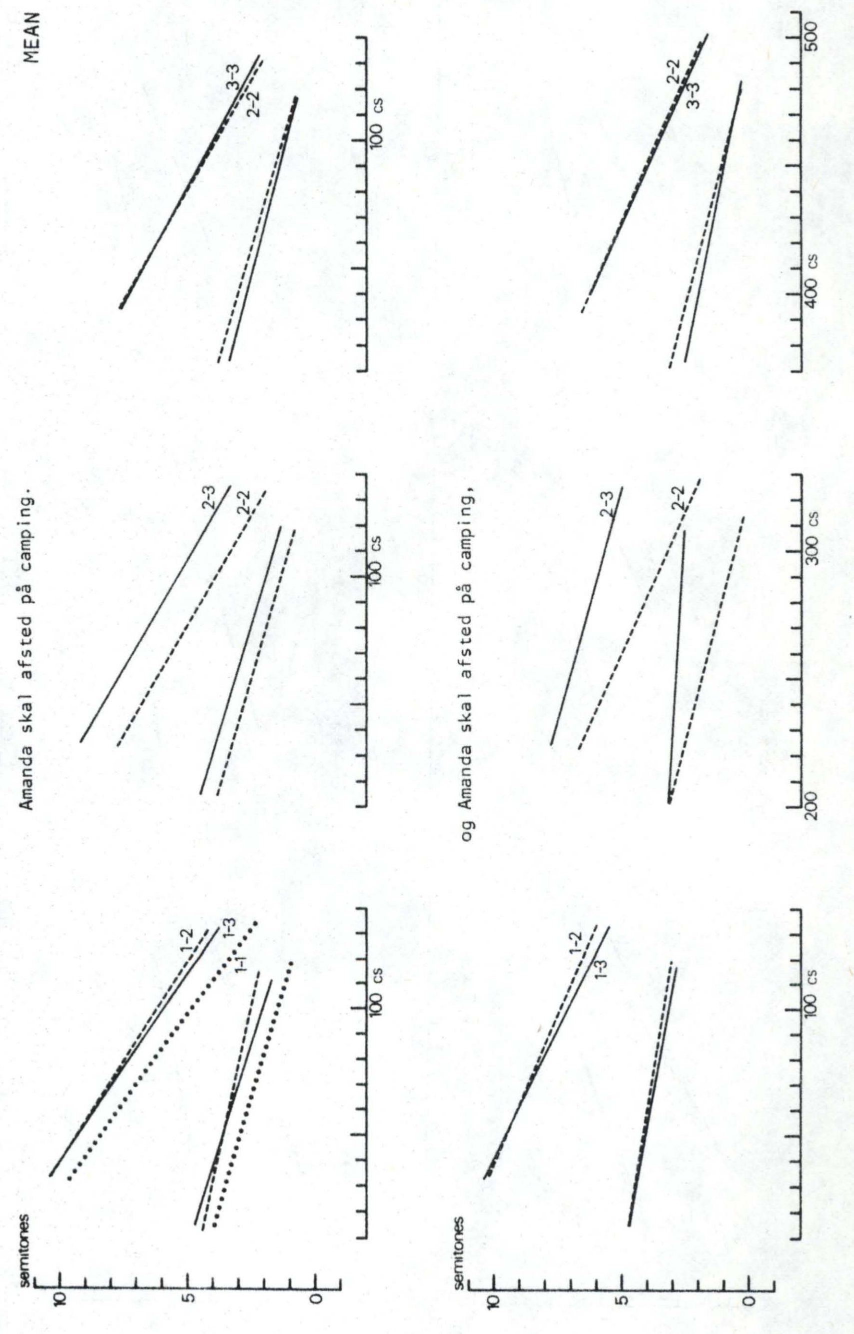

(fig. 2 continued)



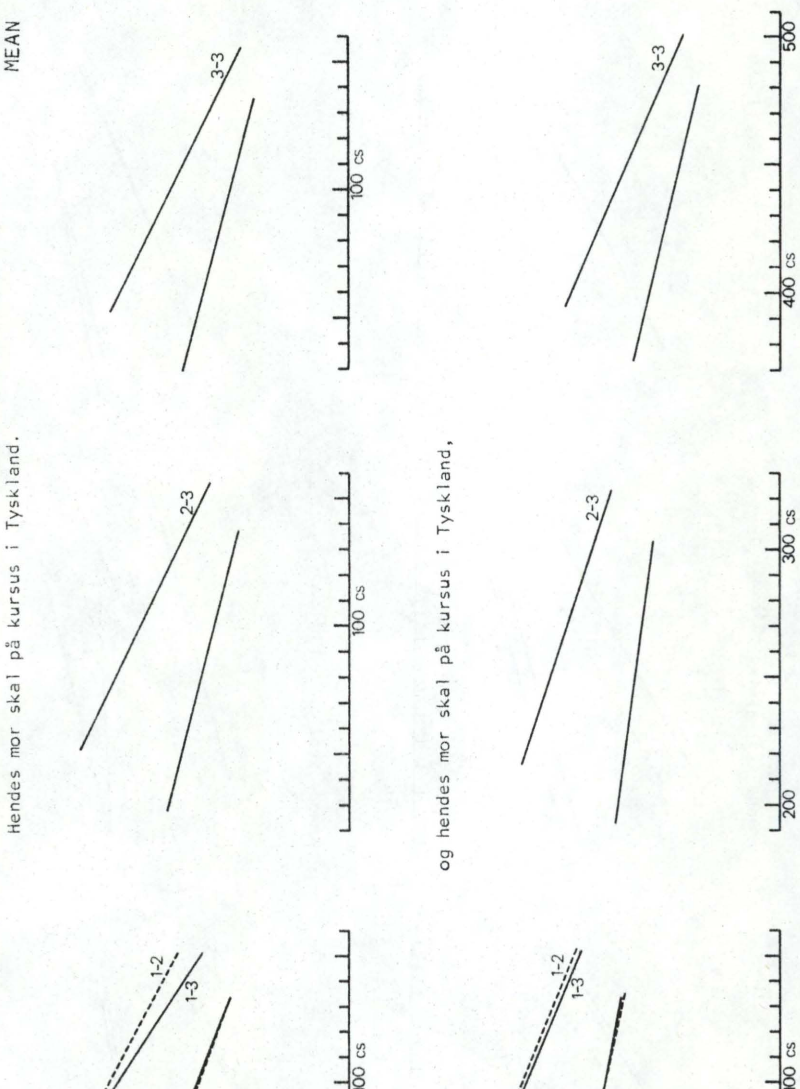

*(fig. 2 continued)*



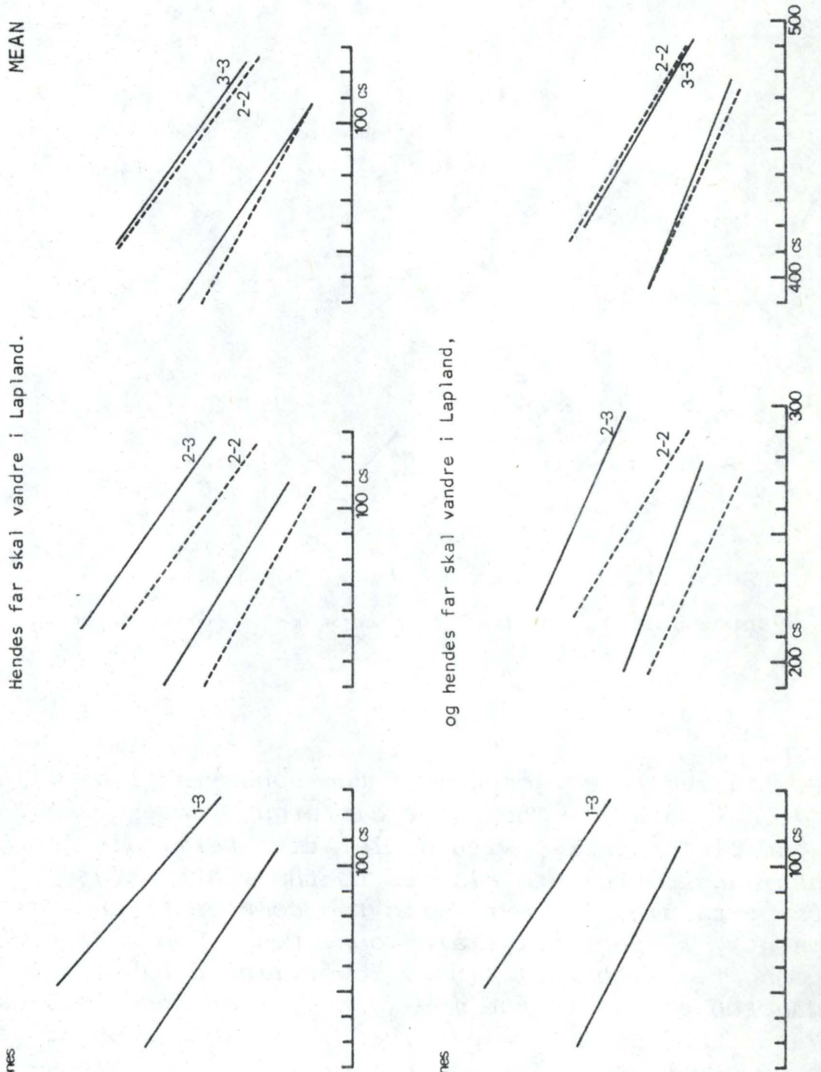

*(fig. 2 continued)*




(fig. 2 continued)




*(fig. 2 continued)*




*(fig. 2 continued)*

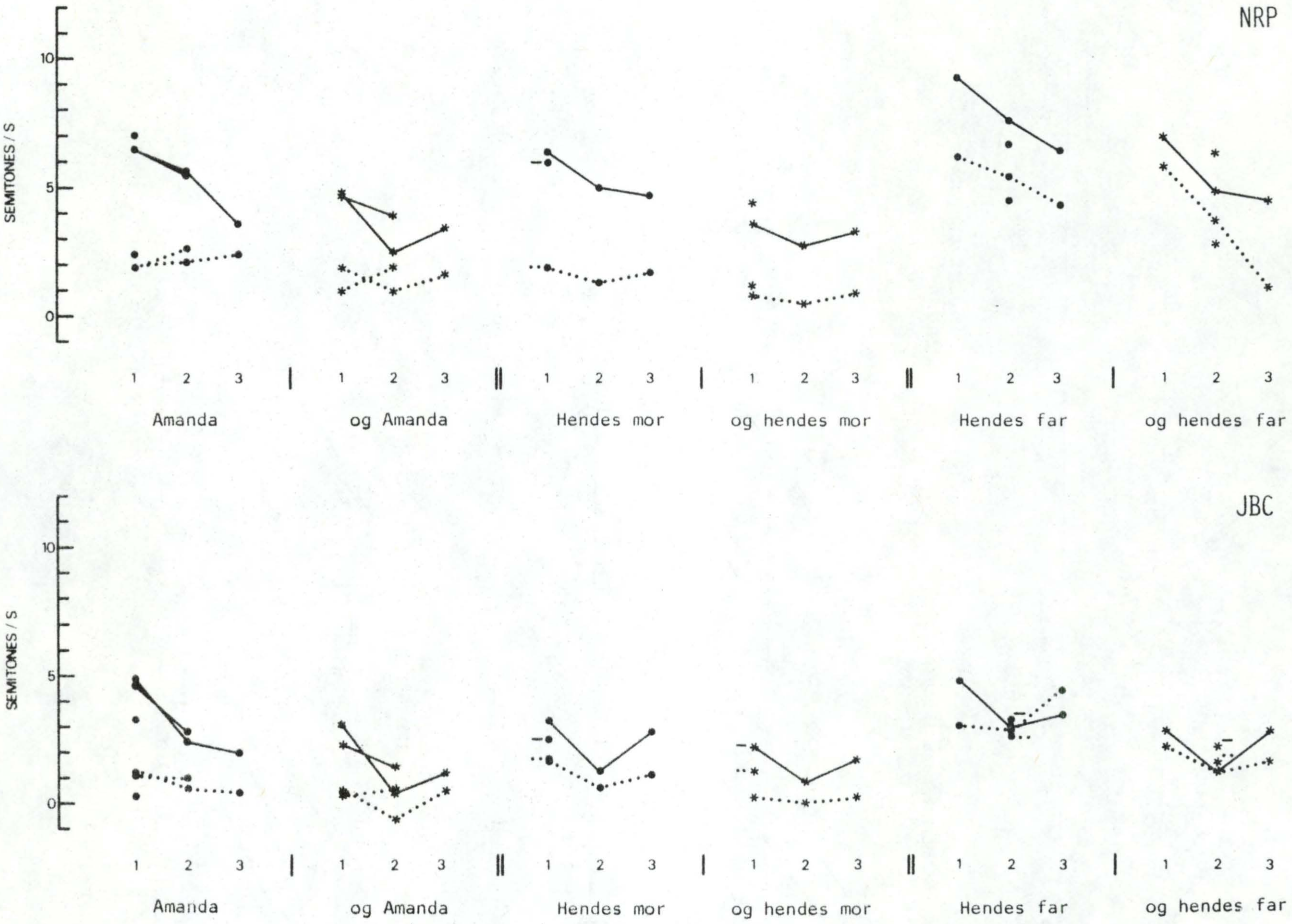
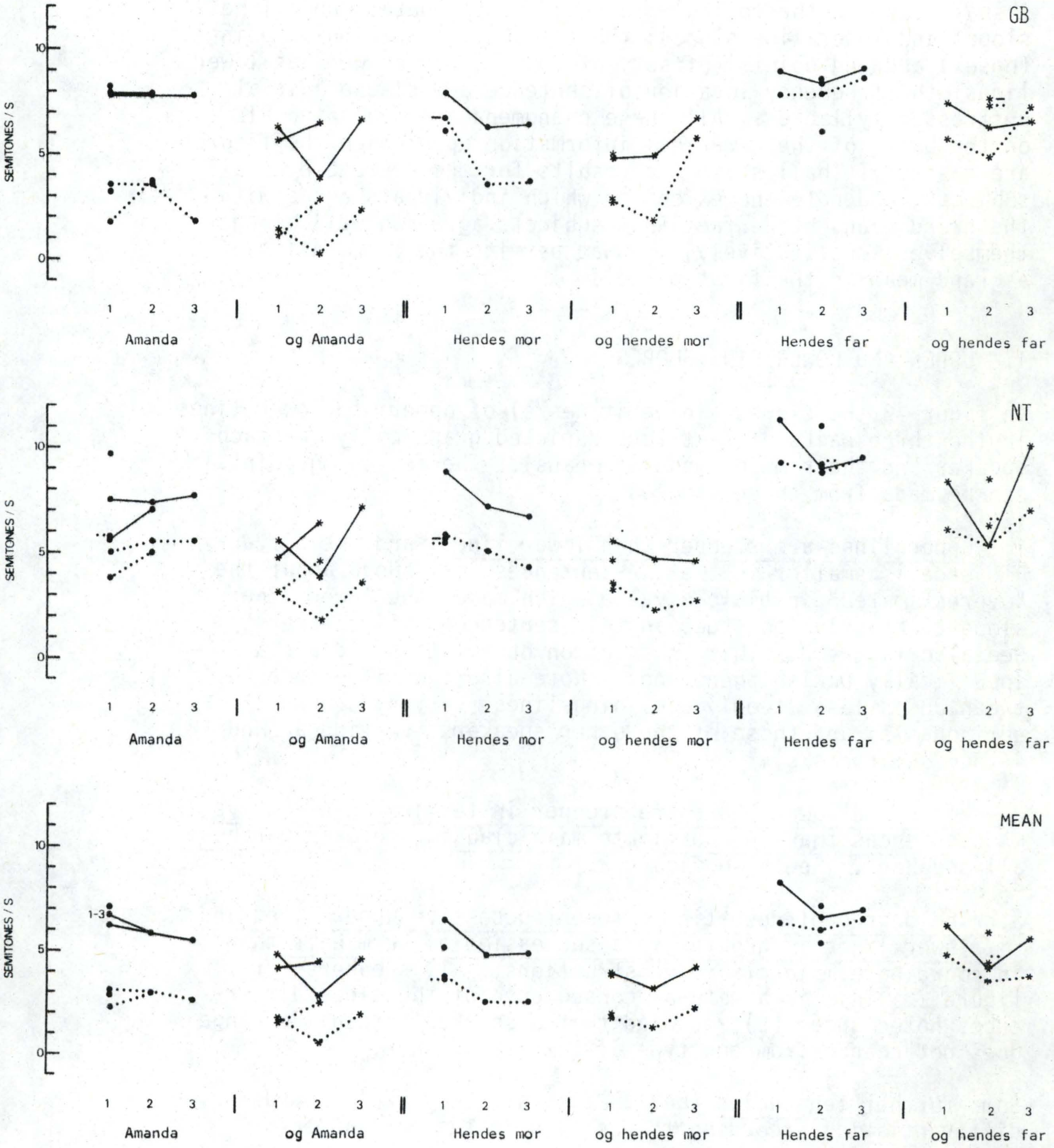



*Figure 3*

*Slopes (in semitones/second) of the upper and lower lines depicted in figure 2. The three different sentences are arranged horizontally in pairs, with declarative terminals to the left (points) and coordinate clauses to the right (stars). Data points pertaining to texts with two components and with three components, respectively, are connected. Upper line data points are connected with full lines, lower line data with dotted lines. Unconnected points in the mór- and fár-sentences represent initial and final items, respectively, in two-component texts. Where unconnected points are not clearly associated - in the graph - with either upper or lower line data, identification is ensured by a short full or dotted line next to the dot or star. Four speakers, two males (NRP and JBC) and two females (GB and NT) and their grand mean (mean of means) - identified at the top right of each graph.*




*(figure 3 continued)*

## B. INTERPRETATION

I shall take up the following aspects of the data: upper line slopes and lower line slopes; the frequency of beginning points (onset) and end points (offset) of upper lines as well as lower lines; the frequency location of sentence and clause initial unstressed syllables. All these phenomena are evaluated also on the basis of the numerical information upon which the figures are based. I shall state the results for the average over all subjects, and note the extent to which individuals agree with the grand mean. (The fact that subjects agree so well among themselves, qualitatively, is what permits the calculation of a grand mean in the first place.)

### 1. UPPER AND LOWER LINE SLOPES

In figure 3 the slopes (in semitones/s) of upper and lower lines in the three pairs of texts are depicted graphically for each speaker and their mean (mean of means). The following points can be made from these figures:

1. Upper lines are steeper than lower lines, and the difference is smaller in the *fár*-sentences, cf. above about the "overestimated" initial Fø value which makes the lower line slope relatively too steep in this sentence: all speakers. See also figure 2. This is a common observation and not a specifically Danish phenomenon. Note also the difference between JBC's less steeply declining lines (varying around 2½ semitones/s) and those of the other speakers (varying around 4 to 6 semitones/s).

2. Upper and lower lines are steeper in terminal declarative sentences than in coordinate main clauses, ceteris paribus: all speakers. See also figure 2.

3. The degree of resetting between successive upper lines and lower lines is greater in a succession of terminals than in coordinate main clause constructions: all speakers, cf. figure 2. This is a natural consequence of the slope differences noted under (2), granted that a speaker's total Fø range does not change from one type of text to another.

Some further tendencies (but that is all they are) towards a differentiation according to position in the text appear:

4. The upper and lower line slopes in an isolated sentence are equal to or very slightly steeper than initially in multi-sentence texts (relevant only for terminal sentences): though with JBC upper and lower lines are less steep in the isolated sentence. Otherwise, upper and lower line slopes tend to be more nearly identical in initial sentences and in final sentences, irrespective of text length. I.e., 1-2 resembles 1-3 and 2-2 resembles 3-3 (rather than 2-3). See also figure 2.

5. Upper and lower line slopes in the terminal sentences tend to be steeper in initial than in non-initial position in three sentence texts: there are random exceptions with all speakers, typically where a final slope is as steep as the initial one. The two sentence text offers no uniform picture. See also figure 2.

6. Upper and lower line slopes in coordinate main clauses are less steep in medial position than initially and finally - where they are approximately equal - in three sentence texts, and slopes are steeper finally than initially in two sentence texts: there are a few slight and random exceptions with every speaker. See also figure 2.

The full and dotted lines in figure 3 run approximately parallel to each other, considered frame by frame, with a couple of apparently random exceptions. In other words, the difference in upper and lower line slope is nearly constant across a given multi-sentence text. I.e., given a certain change in lower line slope across a two- or three-sentence text, the upper lines follow suit, and can be deduced from the behaviour of the lower lines. (Arguments to the effect that the lower line - the stressed syllables of the utterance - is the independent variable are given in e.g. Thorsen 1980b and 1980c. Suffice it here to say that every utterance has a lower line (one or more stressed syllables), but an utterance does not invariably have an upper line which depends for its existence upon the presence of a post-tonic syllable in the prosodic stress group(s).) The steeper upper lines derive from the fact that the magnitude of the Fø interval between the minimum in the stressed syllable and the maximum in the post-tonic varies with position on the intonation contour (early - larger interval, later - smaller interval), cf. Thorsen (1980b, 1984).

From the figures and (1) to (6) above we learn that the individual components are subordinated differently to the superordinate declination which spans the whole text. Individual declinations are steeper and have greater amounts of resetting between them in a series of declarative terminal sentences than in a corresponding sequence of coordinate main clauses. Furthermore, in a series of three terminals the individual slopes tend to become successively less steep, whereas coordinate clauses level out in medial position and a final declination is always steeper than a preceding (initial or medial) one. This can be taken to indicate a more integrated and coherent intonational structure accompanying the closer relation between coordinate main clauses.

Finally, it appears that increasing text length from two to three sentences/clauses or, inversely, decreasing text length from three to two, will leave the initial and final components unaffected. A component is squeezed in or removed medially with no apparent changes in the surroundings. As a result, the second component in a text is different when it is simultaneously text final from when it is medial (compare 2-2 and 2-3 in figure 2). In other words, the production of the second

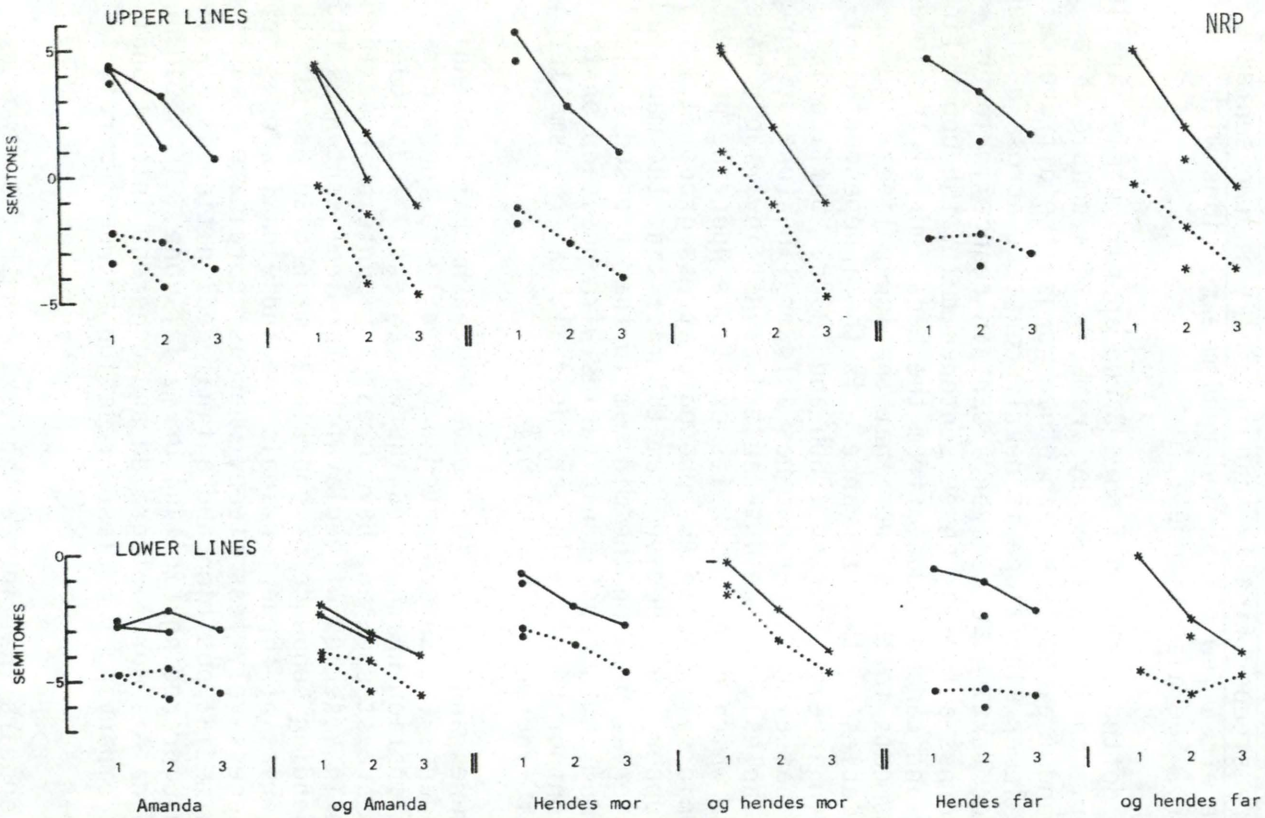
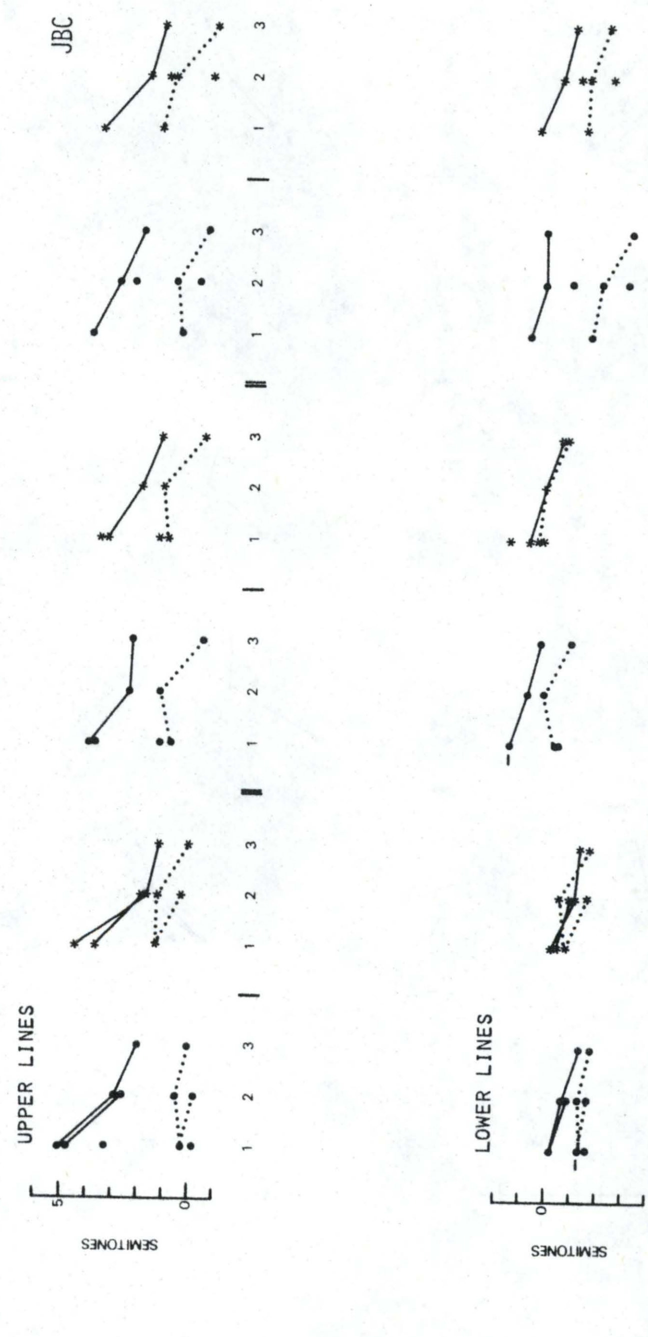
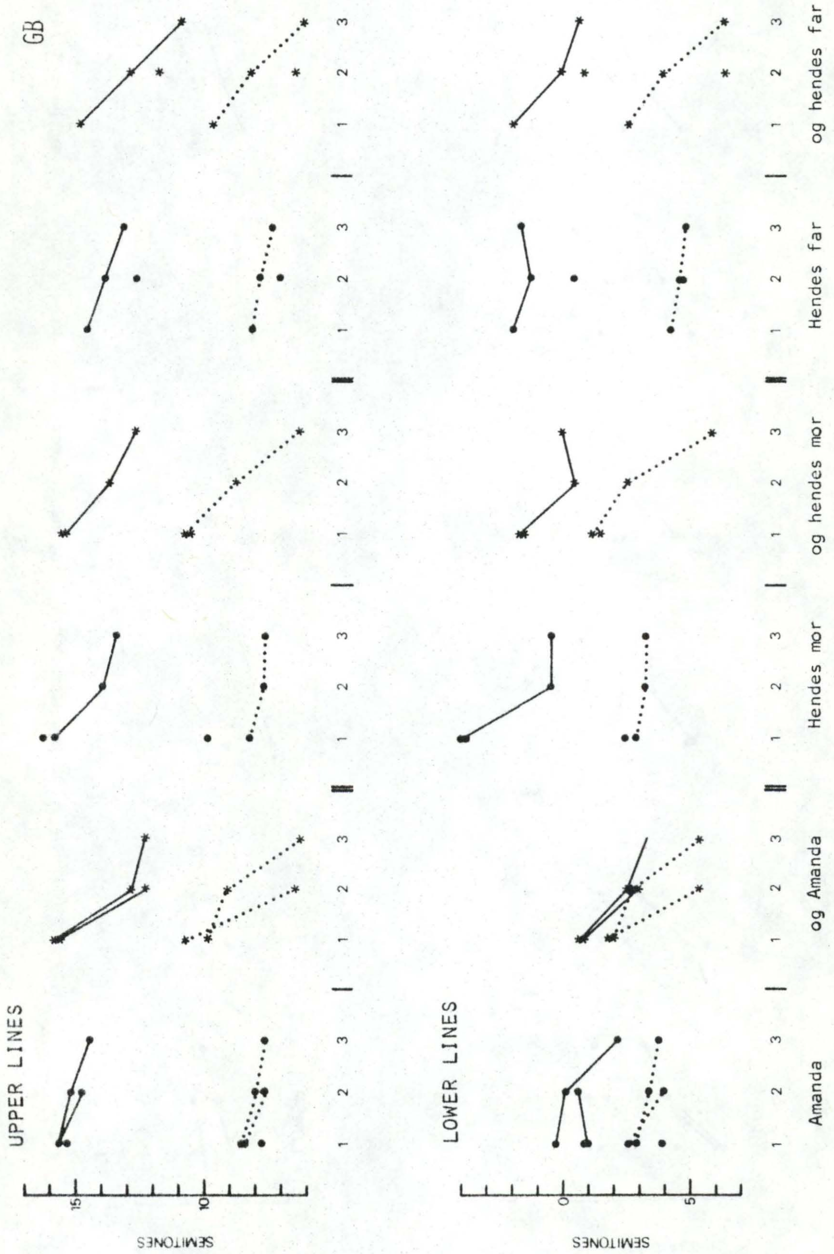



*Figure 4*

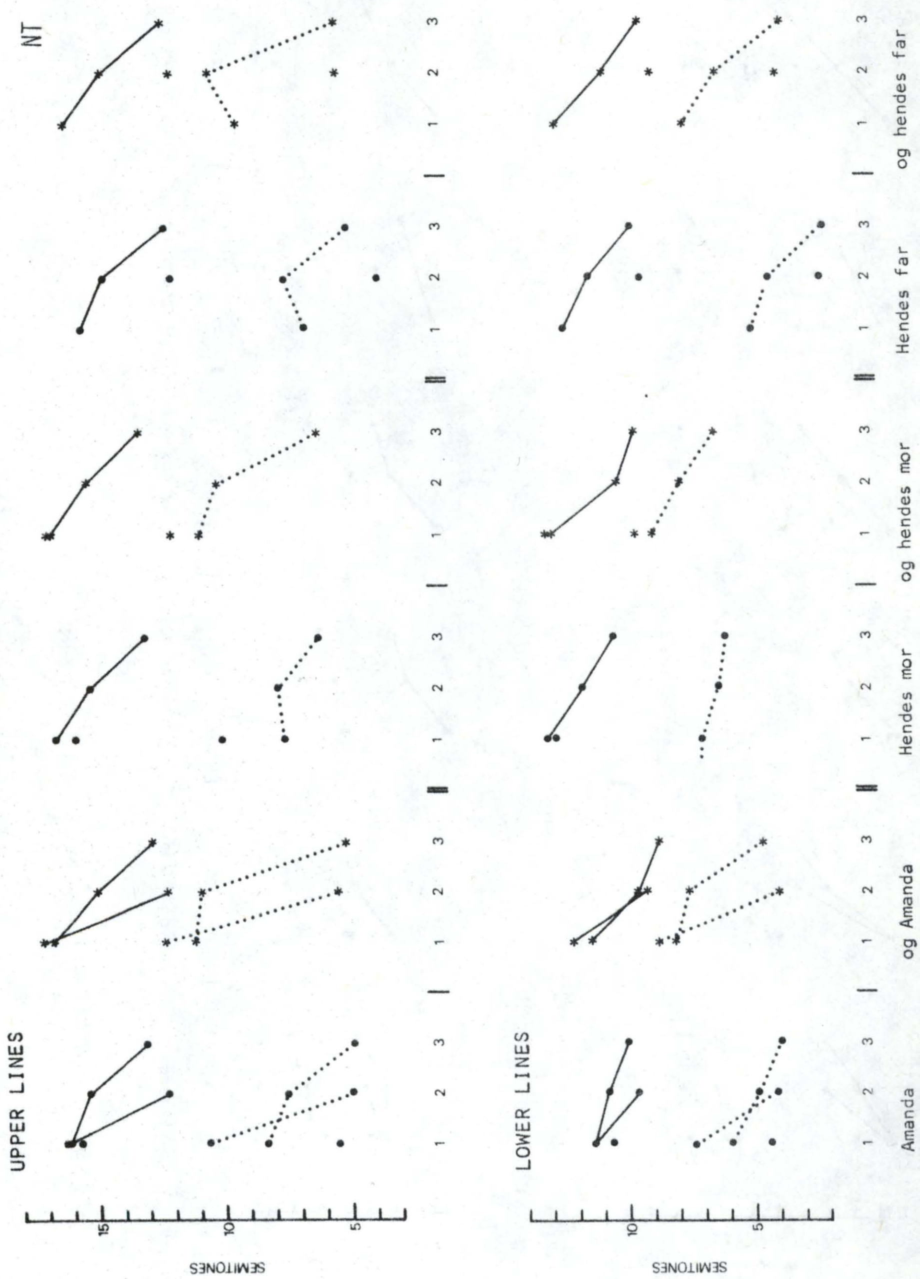
*Frequency values of beginning and end points of the upper (above) and lower (below) lines depicted in figure 2. The three different sentences are arranged horizontally in pairs, with declarative terminals to the left (points) and coordinate clauses to the right (stars). Data points pertaining to texts with two components and with three components, respectively, are connected. Beginning points are connected with full lines, end point data with dotted lines. Unconnected points in the mór- and fár-sentences represent initial and final items, in two-component texts. Where unconnected points are not clearly associated - in the graph - with either beginning or with end point data, identification is ensured by a short full or dotted line next to the dot or star. Zero on the logarithmic frequency scale corresponds to 100 Hz. Four speakers, two males (NRP and JBC) and two females (GB and NT) and their grand mean (mean of means) - identified at the top right of each set of graphs.*



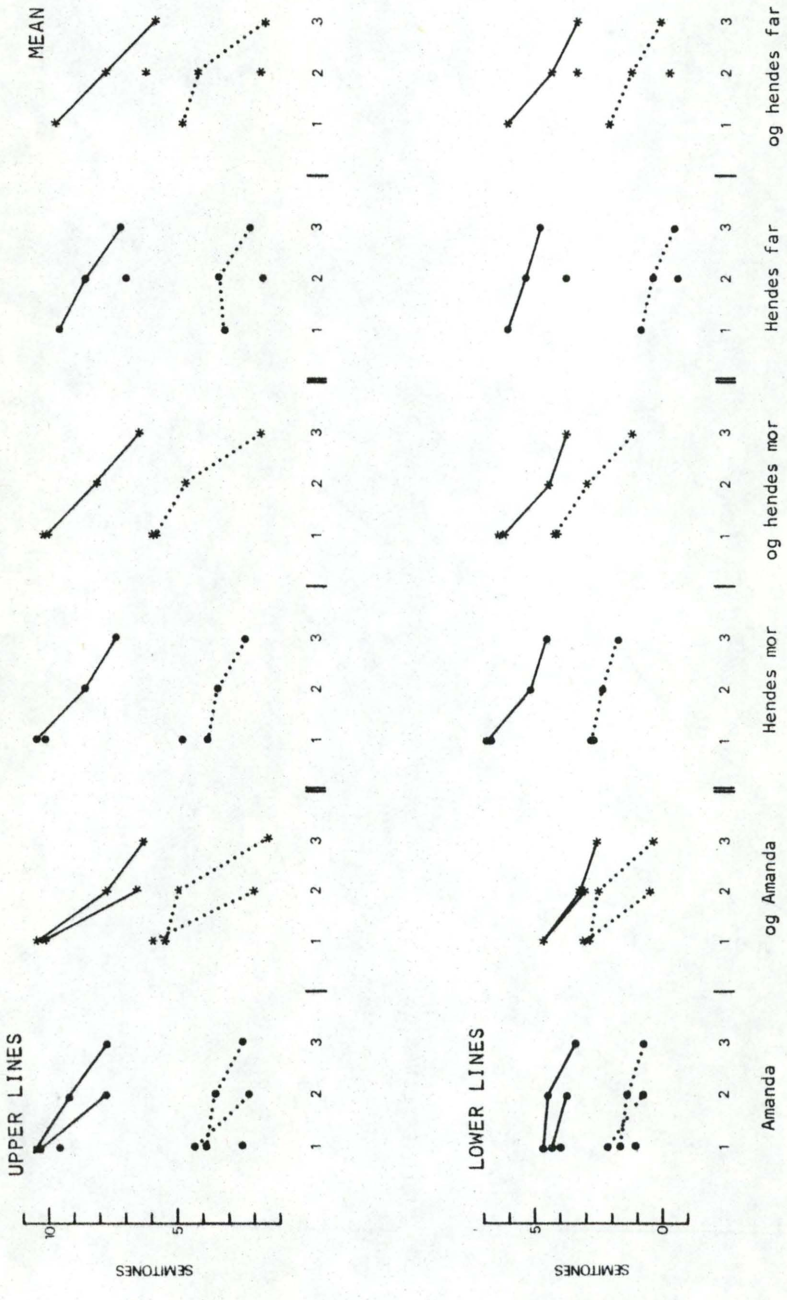

*(fig. 4 continued)*




*(fig. 4 continued)*




(*fig. 4 continued*)




*(fig. 4 continued)*

component is sensitive not only to what precedes it but also to upcoming events. This argument extends to isolated versus text initial sentences, of course. See further section III below.

## 2. UPPER AND LOWER LINE ONSET AND OFFSET

In figure 4 beginning and end point frequencies (full vs. dotted tracings) of upper and lower lines in the three pairs of texts are depicted graphically for each speaker and their grand mean (mean of means). From these figures the following points can be made:

First of all, there is no qualitative difference between texts with terminal sentences and texts with coordinate main clauses but

7. changes in onset and offset frequencies across a text are larger in coordinate main clause constructions:
all speakers, but with NT the difference between terminal sentences and coordinate clauses is small in this respect. The general trend is a consequence of (2) and (3) above: clause contours are less slanted relative to the same overall decline, so onsets and offsets decline more through the text. See also figure 2.

With this difference in mind, "clause" can substitute for "sentence" at any point in the following four statements (8-11):

8. Upper line onsets decrease smoothly across each text:
all speakers, with spurious exceptions to the smoothness by GB and NT. A final upper line onset is 3-4 semitones lower than an initial one, on the average. Upper line onsets are very nearly identical across initial sentences, including the isolated one, irrespective of text length, and across final sentences (excluding the isolated one), i.e. 1-1 resembles 1-2 which resembles 1-3, and 2-2 resembles 3-3: all speakers, except JBC where 1-1 resembles 2-2 (and 3-3). See also figure 2. Initial upper line onsets are also identical across the two boundary conditions: all speakers.

9. Upper line offsets decrease across each text but not quite smoothly, since the drop from 1-3 to 2-3 is smaller than between 2-3 and 3-3: with JBC and NT we even get a "rise-fall", i.e. the end point in 2-3 is higher than in 1-3 but lower in 3-3 than in both preceding ones, most notably in the coordinate clauses, where it is due to the way upper line slopes develop: a steep initial one, then a levelling out medially, and then again a steeper final slope, cf. (6) above. Offsets also decrease by 3-4 semitones (average) from initial to final position. Upper line offsets are very nearly identical across initial sentences (excluding the isolated one), irrespective of text length, and across final sentences, including the isolated one: all speakers. See also figure 2. Final upper line

offsets are identical across the two boundary conditions: all speakers, except GB where they are lower in coordinate main clause texts.

10. Lower line onsets decrease across each text: NRP has one instance of a step up between 1-3 and 2-3, and GB gives no uniform picture. A final lower line onset is 2 semitones lower than an initial one, on the average. In the lower line the onset of an isolated sentence is intermediate between onsets in initial and final sentences in multi-sentence texts (relevant only in the terminals): with NRP and NT it is much closer to text initial lower line onsets, and with JBC it is - conversely - closest to text final onsets. See also figure 2. (If we compare upper and lower lines it is apparent that lower line onsets and offsets span a smaller Fø range than in upper lines, and so the same slightly lower beginning of upper and lower lines in an isolated sentence compared with initial sentences in multi-sentence texts will bring isolated lower line onsets relatively nearer to the onset of lower lines in final sentences.) Otherwise, lower line onsets are very nearly identical across initial sentences (i.e. 1-2 equals 1-3) and across final sentences (i.e. 2-2 equals 3-3): all speakers, with a couple of spurious exceptions. See also figure 2. Initial lower line onsets are identical across the two boundary conditions: all speakers, except GB's *mór*-sentences.

11. Lower line offsets decrease across each text: no speaker is as regular as the grand mean. As with upper lines, lower line offsets are very nearly identical across initial sentences (excluding the isolated one), irrespective of text length, and across final sentences, including the isolated one: with NRP the isolated sentence resembles initial rather than final sentences. See also figure 2. Final lower line offsets are also identical across the two boundary conditions: all speakers, except GB where they are lower in final coordinate clauses (as are her upper line offsets, cf. (9) above).

With a reservation about the onset of the lower line in an isolated sentence (cf. (10)), the ensemble of upper and lower lines in a text span almost the same Fø range, irrespective of text length and boundary condition. The upper and lower lines, respectively, begin at the same frequency in text initial sentences/clauses and end at the same frequency in text final sentences/clauses. Thus, the superordinate declination varies with text length but it is the same in both boundary conditions. The fact that both onsets and offsets of upper and lower lines are identical in initial sentences (1-2 resembling 1-3) and in final sentences (2-2 resembling 3-3) ties up with the conclusion drawn from points (1)-(6) above. The second component in a text is different when it is simultaneously text final than when it is medial, and the same is true of an isolated versus a text initial sentence. Together the upper and lower lines bear testimony to a considerable amount (in terms of temporal scope) of preplanning and look-ahead in the production of intonational events. It is possible - though it remains to be tested - that only three different textual positions are distinguished, as far as sentence intonation onset and slope are concerned: initial, final, and medial. That is to say, two or more medial sentences or clauses may not be further differentiated among themselves. Such an arrangement of individual text components would be less taxing on the speaker's pre-planning operations (cf. section III).

## 3. INITIAL UNSTRESSED SYLLABLES

In text initial position the *Amánda*-sentences have only one unstressed syllable, [a], the *mór*- and *fár*-sentences have two (the second of which may be carried by a syllabic nasal as a result of schwa-elision) but only the first one, [hen], will be dealt with here. In text medial position an unstressed *og*, [ʌ], is added to the beginning of each coordinate main clause. This syllable, [ʌ], is - with very few exceptions - higher than the succeeding [a] or [e], and by nearly the same amount for each subject (0.5 to 1.0 semitone with NRP and JBC; 1.0 to 2.0 semitones with GB and NT). Therefore, the conjunction can be derived from [a] and [e] and is not considered any further here.

In figure 5a,b,c the frequency of sentence or clause initial unstressed [a] *(Amánda)* and [e] *(hendes mór, hendes fár)* in initial sentences (5a), final sentences (5b) and across a three sentence text (5c) is depicted, for each speaker and speakers' grand mean (mean of means). Again, I shall state the results for the mean and note the extent to which each subject agrees with the overall tendencies.

12. The pretonic syllable(s) in text initial sentences are constant across texts of different length and different syntactic conditions: all subjects, with exceptions by GB, cf. figure 5a. This (near) constancy in text initial unstressed syllables is analogous to the constant onset of lower lines (i.e. to the constant frequency of the first stressed syllable) across text lengths and boundary conditions, cf. (10) above.

13. Sentence initial unstressed syllables in text final sentences are constant across different text lengths in terminal sentence texts: there are spurious exceptions by NRP, GB and NT, cf. figure 5b. This corresponds to the behaviour of the final lower line onset, i.e. the frequency of the first stressed syllable (cf. (10)). In coordinate main clause constructions the pretonic syllables in the last clause tend to be lower in the longer text: JBC is an exception; his final pretonics are constant in this condition as well. The general tendency is likely to be due to a higher offset of the declinations preceding 2-2 than the ones preceding 3-3, i.e. the final Fø minimum and the (derived) peak in the preceding clause is higher in the first clause of two (1-2) than in the second of three (2-3), see figure 4 ("lower lines"-data points connected by dotted line) and also figure 2.

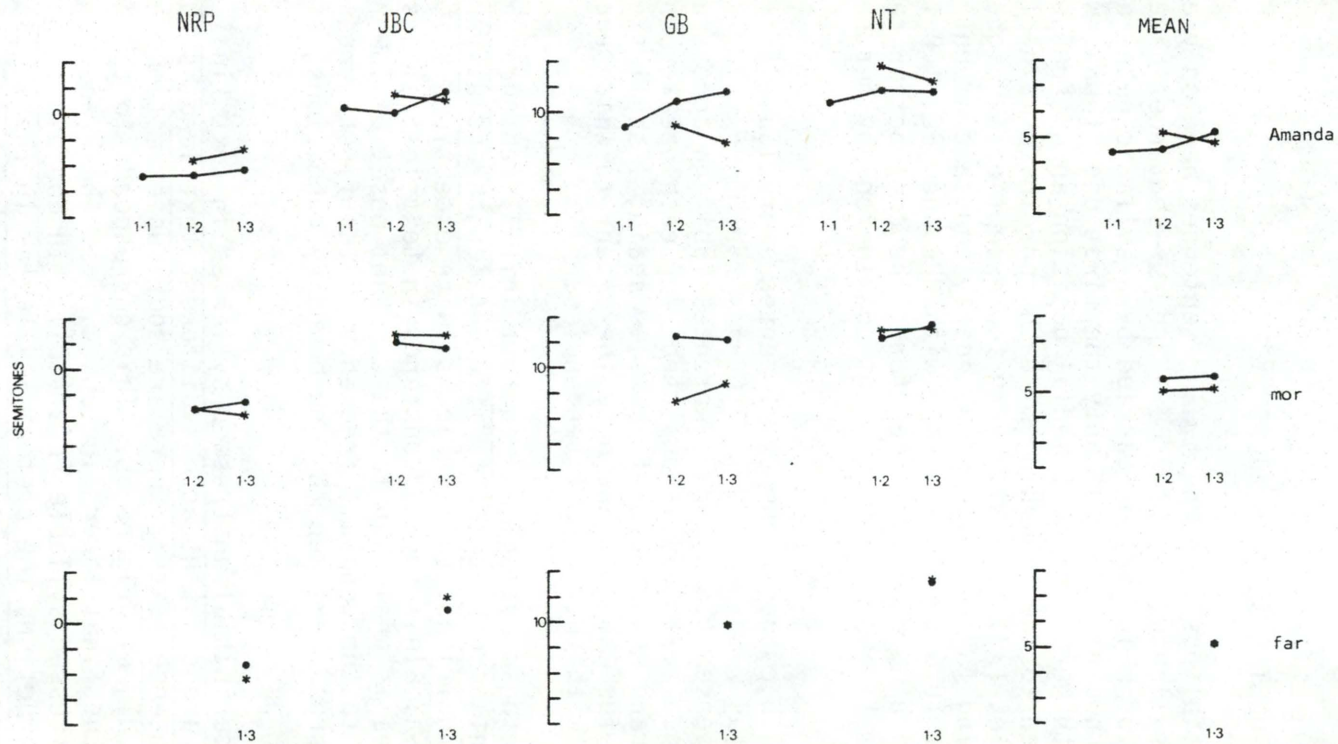



*Figure 5a*

*Frequency values of sentence or clause initial unstressed syllables in the sentences whose upper and lower lines are depicted in figure 2. The three different sentences are arranged vertically, as indicated by key words to the right of each graph. Data pertaining to declarative terminal sentences are indicated with points, coordinate clause data with stars. Data from text initial sentences (including the isolated one) and clauses are plotted in figure 5a, data from text final sentences (including the isolated one) and clauses are plotted in figure 5b, and data from initial, medial and final items in three component texts are plotted in figure 5c. Zero on the logarithmic frequency scale corresponds to 100 Hz. Four speakers, two males (NRP and JBC) and two females (GB and NT) and their grand mean (mean of means) - identified at the top of each figure.*

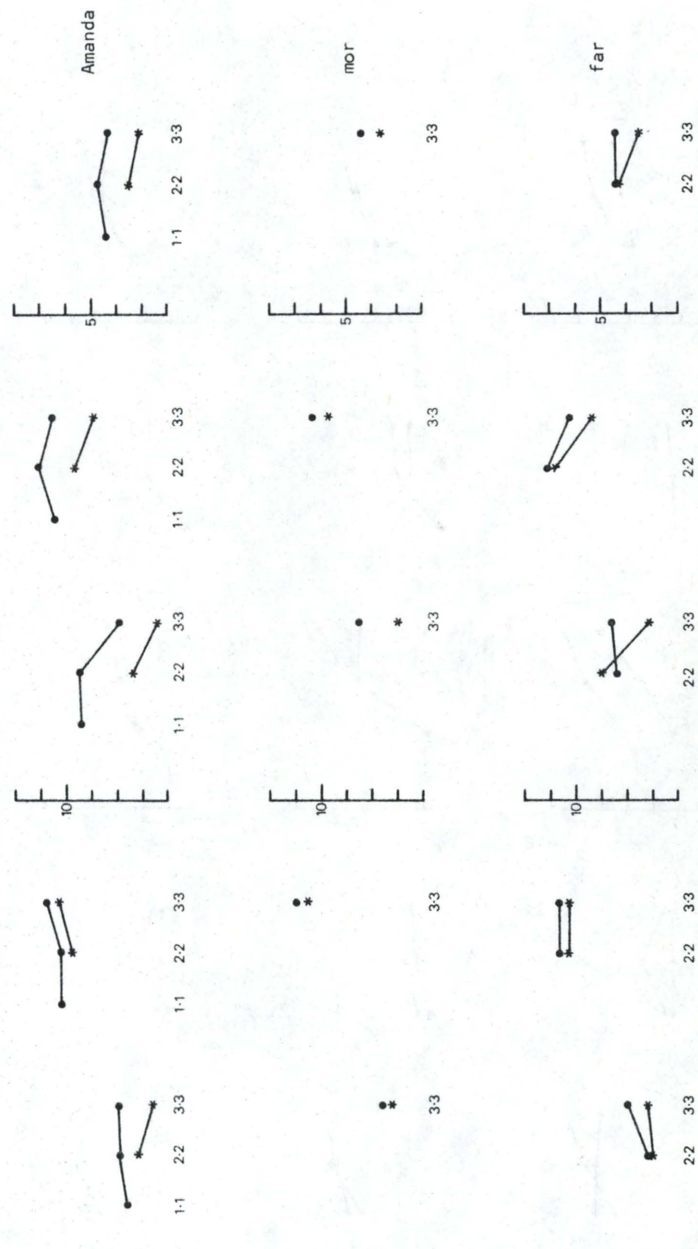



*Figure 5b*

*See the legend to figure 5a*

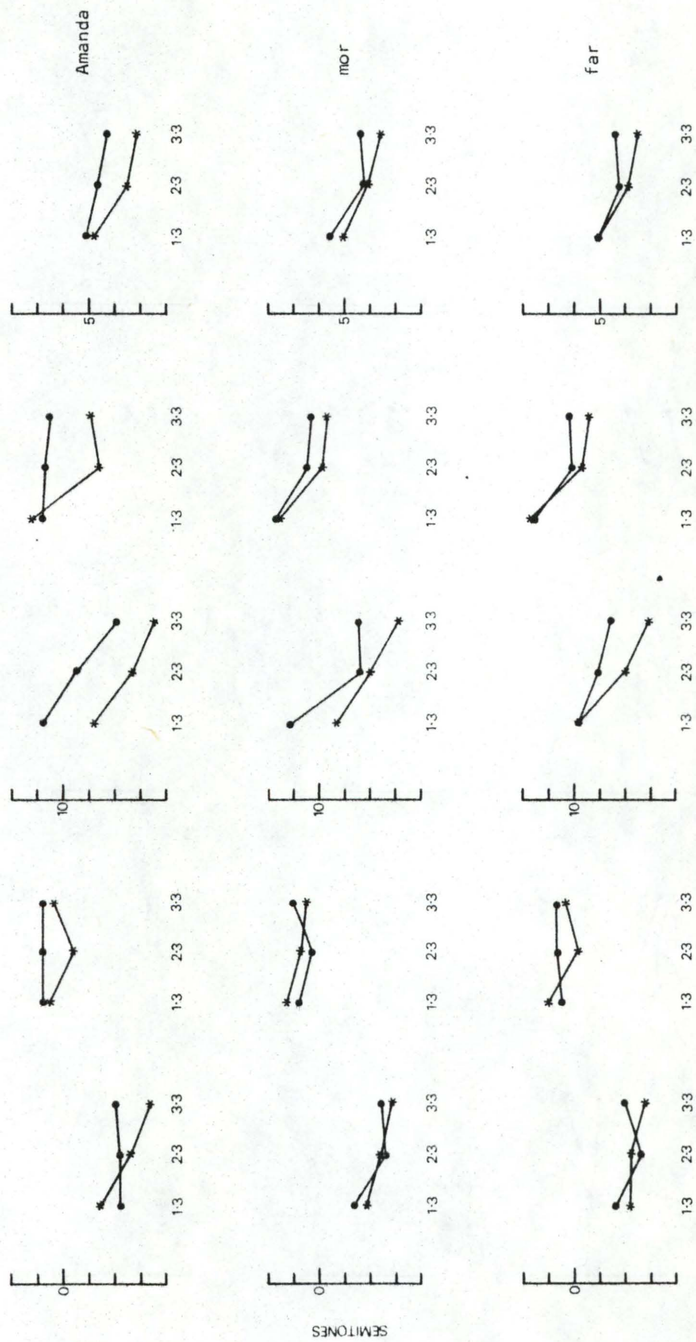



*Figure 5c*

*See the legend to figure 5a*

At least this difference is greater, and corresponds better to the difference observed in the clause initial unstressed syllables in figure 5b than does the slight or even reverse differences in the onset of lower lines in 2-2 and 3-3. This would mean that clause initial unstressed syllables in text medial position are governed tonally by the offset of the preceding clause, or - in other words - such syllables join up prosodically with the last stress group in the preceding clause. They are neither isolated tonally, nor do they attach themselves to the clause to which they belong syntactically. This interpretation gets a certain substantiation from figure 1 (and from the corresponding longer texts not reproduced here): there are a majority of instances (exceptionless with GB and nearly so with NT) where clause initial unstressed syllables perform a smooth and continuous fall together with the posttonic in the preceding clause; together with the last stressed syllable in the preceding clause the ensemble of unstressed syllables thus perform the low plus high-falling Fø pattern characteristic of prosodic stress groups in Standard Danish (cf. Thorsen 1980b, 1984). Only with JBC are there no such cases of prosodic association across clause boundaries, maybe because he (as the only one) paused between the coordinate clauses. (I do not think that a pause necessarily disrupts a tonal pattern (apart from the introduction of a silent interval), and clause boundaries per se apparently do not, but the combination: pause and clause boundary may be more than an Fø pattern can survive.)

14. Initial unstressed syllables through a multi-sentence text decrease slightly through a string of terminal sentences: the variation is small and non-uniform with NRP, nearly nil with JBC, apparent only in NT's *mór*- and *fár*-sentences. With GB the decrease is considerable in the *Amánda*- and *mór*-sentences, cf. figure 5c. Again, this is analogous to the onset of lower lines, i.e. to the behaviour of the first stressed syllable in each clause in the texts, cf. (10) above. In coordinate main clause texts the decrease is greater and generally more uniform: all speakers. Again, this is more in line with the way upper line offsets develop through a text than with lower line onsets (cf. figure 4), i.e. it seems to be caused by the tonal association of the clause initial unstressed syllables with the preceding clause final stress group, cf. above.

What points (12)-(14) demonstrate is merely a tonal dependency of initial unstressed syllables in the sentence or clause upon other parameters: In text initial position, initial unstressed syllables are constant (for a given speaker), as is the first stressed syllable, across different text lengths and different boundary conditions. After a text medial sentence boundary, initial unstressed syllables likewise associate with the succeeding stressed syllable, i.e. differences in frequency location across different text positions are analogous with and can be derived from differences in the first stressed syllable in the sentence. After a clause boundary (unaccompanied by any pause) clause initial unstressed syllables (often but not invariably) associate tonally with the preceding prosodic stress group, i.e. differences across different text positions are analogous with the differences in the last tonal pattern in the preceding clause. In other words, sentence and clause initial unstressed syllables in text medial position may be extrametrical phonologically (and rhythmically as well, after a sentence boundary), but tonally they are not isolated entities; they vary in a principled fashion in association either with a preceding or a succeeding prosodic stress group.

## C. DURATION

I have measured the duration of each sentence and clause, as the time interval between the first and last Fø measuring point, excluding the conjunction in the coordinate clauses. Average values for each speaker, and speakers' mean are given in Table I.

The differences in duration of a given sentence across text positions - in one or the other boundary condition - range between 16 cs (NRP's coordinate *fár*-clauses) and 3 cs (NRP's terminal *fár*-sentences), i.e. 13.7 and 2.5%, respectively, of the shortest sentence/clause. The most consistent trend in the durations is that an initial sentence or clause is shorter than the succeeding, medial or final one (true in 29 out of 32 instances). Secondly, a final sentence or clause tends to be shorter than a medial one in three sentence texts (true in 16 out of 24 instances). Isolated sentences are not systematically different from sentences in context: though with JBC the isolated sentence is clearly longer than initial sentences (6 and 7 cs), it is only 2 or 3 cs longer than medial and final sentences. With GB the isolated sentence is 7 cs longer than the final sentence of three but only 1 cs longer than initial sentences. Lehiste (1975) found that isolated sentences were longer than in three sentence texts. To judge from the present data, such a finding may not generalize to all speakers. Furthermore, coordinate main clauses tend unexpectedly to be longer (by around 6 cs) than terminal sentences (which is not due to the conjunction *og* which was excluded from the measurement). The opposite relation would perhaps have been less surprising: A terminal sentence develops freely, a non-terminal clause is shortened by the succeeding clause. I have no explanation to offer at the moment for the longer coordinate clauses.

A more thorough treatment of durational differences warrants a separate investigation. The only conclusion I wish to draw at present is that sentences or clauses are shorter text initially than in later positions but - somewhat surprisingly maybe - isolated and text final sentences or clauses are not systematically longer than sentences in other positions in the text.

*Table I*

*Average durations in centiseconds of three different sentences in various contexts: isolation ("1-1"), initially ("1- "), medially ("2- "), and finally ("3- ") in texts consisting of two (" -2") or three (" -3") sentences, and in two different boundary conditions, as indicated to the left. See further the text.*

| | NRP | JBC | GB | NT | mean |
|---|---|---|---|---|---|
| *Amánda* | | | | | |
| terminal sentences | | | | | |
| 1-1 | 128 | 135 | 119 | 137 | 129.8 |
| 1-2 | 128 | 128 | 118 | 131 | 126.3 |
| 2-2 | 130 | 132 | 117 | 136 | 128.8 |
| 1-3 | 131 | 129 | 118 | 136 | 128.5 |
| 2-3 | 132 | 133 | 118 | 137 | 130.0 |
| 3-3 | 133 | 133 | 112 | 143 | 130.3 |
| coordinate clauses | | | | | |
| 1-2 | 131 | 136 | 114 | 136 | 129.3 |
| 2-2 | 139 | 146 | 122 | 141 | 137.0 |
| 1-3 | 129 | 136 | 120 | 134 | 129.8 |
| 2-3 | 134 | 141 | 125 | 141 | 135.3 |
| 3-3 | 136 | 135 | 120 | 137 | 132.0 |
| *Hendes mór ...* | | | | | |
| terminal sentences | | | | | |
| 1-2 | 148 | 157 | 140 | 145 | 147.5 |
| 1-3 | 148 | 153 | 139 | 146 | 146.5 |
| 2-3 | 156 | 157 | 144 | 149 | 151.5 |
| 3-3 | 152 | 157 | 145 | 148 | 150.5 |
| coordinate clauses | | | | | |
| 1-2 | 151 | 154 | 141 | 150 | 149.0 |
| 1-3 | 152 | 153 | 140 | 150 | 148.8 |
| 2-3 | 159 | 160 | 145 | 156 | 155.0 |
| 3-3 | 159 | 160 | 141 | 153 | 153.3 |
| *Hendes fár ...* | | | | | |
| terminal sentences | | | | | |
| 2-2 | 121 | 138 | 111 | 124 | 123.5 |
| 1-3 | 121 | 128 | 118 | 127 | 123.5 |
| 2-3 | 120 | 132 | 116 | 131 | 124.8 |
| 3-3 | 118 | 129 | 109 | 126 | 120.5 |
| coordinate clauses | | | | | |
| 2-2 | 122 | 144 | 118 | 128 | 128.0 |
| 1-3 | 117 | 133 | 114 | 130 | 123.5 |
| 2-3 | 133 | 135 | 124 | 134 | 131.5 |
| 3-3 | 123 | 130 | 121 | 131 | 126.3 |

## D. SUMMARY OF THE RESULTS ON INTONATION

A short text, composed of two or three simple declarative sentences or corresponding clauses, of which none is excessively long, will be associated with a superordinate Fø declination which supports individually slanting sentence/clause components. The most interesting and consistent further findings, i.e. where the inter subject agreement is best, are points (2, 3) and (8 - 11): Upper and lower lines - which characterize individual sentences or clauses - are steeper and with greater amounts of resetting between them in a succession of declarative terminal sentences than in a corresponding succession of coordinate main clauses. Concomitantly, the onset and offset Fø values of these upper and lower lines decrease across a text, more so through coordinate main clauses, because individual lines are less slanted relative to the same superordinate decline. The decrease is greater in upper than in lower lines, ceteris paribus, because upper lines - individually and as a whole - span a greater Fø range than do lower lines. These facts are depicted in the stylized graph in figure 6, which shows the course of upper and lower lines in two different three-sentence texts, one with declarative terminal sentences (full lines), and one with corresponding coordinate main clauses (dashed lines). (Differences in duration and pause length are obliterated.) The dotted lines reflect the overall downdrift which is identical in the two boundary conditions. They connect the (upper and lower) onset values in

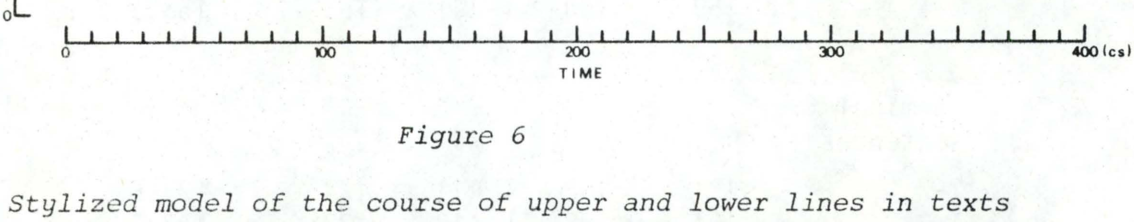



*Figure 6*

*Stylized model of the course of upper and lower lines in texts with three declarative terminal sentences (full lines) and three coordinate main clauses (dashed lines). The dotted lines reflect the overall downdrift through the text, which is identical in the two boundary (sentence or clause) conditions. A text with two components is obtained by leaving out the middle section in the figure and moving together the initial and final components across the gap. An isolated sentence results if the dotted lines are suitably compressed in time.*

the first sentence/clause with the (upper and lower) offset values of the last one. These values can be considered constant for texts of different length (and with different internal (sentence or clause) boundaries), which is another important aspect of the results. An isolated sentence results if these dotted lines are suitably compressed in time. A text with two sentences or clauses is obtained by leaving out the middle section of figure 6 and moving together the initial and final components across the gap. Since the frequency span is (assumed to be) constant, an isolated declarative sentence will decline more steeply than the ensemble of two or more sentences in a text: The results actually do show that isolated slopes are at least as steep as the steepest (initial) sentence in multi-sentence texts. Text initial unstressed syllables are likewise constant across different texts, whereas text medial, sentence or clause initial, unstressed syllables vary in a principled fashion in association with a succeeding and preceding, respectively, prosodic stress group.

It should be evident from the account of the results that the description is true of the overall, average picture exhibited by the four speakers in this investigation. If either one of the four speakers had been the only one in the experiment, the conclusions would have been slightly different on some points, because a certain amount of inter and intra individual variation or "noise" exists. However, this variation is random, i.e. one speaker may differ in one direction from the main trend in one aspect, and in another direction in another aspect; or the same speaker may deviate on a certain point from the main trend in only one of the three different sentences recorded here. Therefore, the presentation of a grand mean over four speakers is possible, on the one hand, while at the same time we can assert that no individual is as regular in his or her behaviour as the resulting model (figure 6) would predict. Speech production is a more plastic process than such a model can capture, also where the intonational structuring of multi-sentence/clause texts are concerned (cf. Thorsen 1984, where I reach a similar conclusion about the production of Fø patterns in prosodic stress groups).

The less steeply declining upper and lower line slopes in coordinate clauses vis-a-vis terminal sentences, ceteris paribus, corresponds well with the observations in Thorsen (1978) that isolated terminal declarative sentences have steeper sentence intonation contours than corresponding non-terminal (declarative and interrogative) clauses.

The fact that different text internal boundaries (clause boundary or sentence boundary) induce differences in intonational manifestation is not a reflection of a general, tight syntactic/prosodic interplay, cf. Thorsen (1980a) - which treats word boundaries and Fø patterning - and Thorsen (1983a) - where long simple declarative isolated sentences are analysed and the relation between intonation contour resettings and sentence internal syntactic boundaries are discussed.

## E. COMPARISON WITH OTHER LANGUAGES

First of all, the present results - like Lehiste's (1975) and Bruce's (1982) - clearly contradict Nakatani's conclusion that speech features [including prosodic ones, NT] can be reasonably generalized to sentences in a coherent context. The upper and lower line(s) of a terminal declarative sentence span a greater frequency range in isolation than in a larger text context. Secondly, Standard Danish agrees well with the gross text intonation features of English and Swedish. All three languages (and Japanese as well, cf. Uyeno et al., 1979) exhibit a superordinate text intonation structure upon which are superposed individual sentence or clause components. There are a number of (minor) differences, however. I cannot assert, as did Lehiste (1975) that an isolated sentence is longer than the same sentence in context. Lehiste further found that Fø in the beginning of a sentence is lower in an isolated one than in paragraph initial position. This is not true here if we consider the very first unstressed syllables, which have a constant Fø value independently of the length of the succeeding text. If we consider the Fø value of the first stressed syllable and especially the first succeeding Fø peak (in the first posttonic syllable), then Lehiste's finding is true of one speaker but the isolated sentence is only slightly lower (less than one semitone) than paragraph initial ones with three speakers. Lehiste's observation that Fø in the beginning of a sentence is higher in paragraph initial than in medial and final position holds for Standard Danish as well, no matter what Fø point we consider to be the beginning, and this difference is not negligible.

Bruce's (1982) observations from his larger material, that there are initial and final Fø values which are constant across texts of different length, hold true of Standard Danish as well, though these constant points pertain to different units in the chain of syllables in the two languages. In Southern Swedish the very first and very last syllable lie at the floor of the range of normal Fø variation. But the first local Fø maximum and minimum (in the first stressed plus unstressed syllable) in Swedish is lower in an isolated sentence than in context (corresponding to Lehiste's (1975) observation), and it is higher in a longer than in a shorter text. The last local Fø maximum and minimum is constant across texts of different length. In the present material, not only extrametrical sentence initial unstressed syllables but the frequency location of the whole of the first prosodic stress group is almost insensitive to text length. Thus, the total range of variation varies proportionately (upwards) with text length in Southern Swedish, whereas it is constant in Danish. Concomitantly, superordinate text declination will vary more in Danish than in Swedish with the duration of the text.

Before the text internal pauses in Bruce's recordings Fø drops to the floor of the range (where also absolute initial and final syllables are constantly located). Such a marking of text medial boundaries is not replicated here. This may have to do with the tonal relation between stressed and unstressed syllables in Standard Danish: the first post-tonic syllable rises above the preceding stressed one, succeeding post-tonics describe a more or less steep fall. In the present sentences there was only one sentence/clause final post-tonic, and a drop to the floor of the Fø range is therefore not possible. The high sentence final fall in Bruce's material may, however, also be due to a sentence accent, which evokes a particularly elaborate Fø movement in the posttonic syllables, and Danish lacks such a nuclear accent (cf. Thorsen 1980b). According to a personal communication by Gösta Bruce, Southern Swedish may also be produced without a final focal accent, but it is not clear to me whether this was in fact the case in his (1982) recordings.

In Bruce's texts, the Fø minimum in the first accent after a text medial sentence boundary continues at the same Fø value that the Fø minimum had in the last accent in the preceding sentence. He sees that as a tonal coupling or assimilation of the succeeding constituent sentence to the preceding one, see further below, section III. I am inclined to think, though, that this relation is an artifact of his material, due to the particular length of his sentences (two stressed words in each). With longer sentences he might have found the onset of a succeeding sentence to be higher than the offset of the preceding one, as it is in the present Standard Danish material, cf. figure 6. Incidentally, the Danish data confirm Cooper and Sorensen's (1981) observation, that the resetting between clauses in a complex sentence is independent of any accompanying pause.

Uyeno et al. (1979) found no difference between coordinate main clause constructions and successions of simple terminal sentences in Japanese. In Standard Danish, the closer relation between coordinate main clauses is accompanied by a more integrated and coherent intonational structure than a series of terminal sentences.

## IV. ON THE ABSTRACT REPRESENTATION OF INTONATION

By "abstract representation" I mean the level where significant and distinctive choices are made and from which the phonetic implementation can be derived by rules. It contains <u>no</u> redundant specifications. This representation may correspond to a stage (early) in the actual speech production process. Thus, a representation of intonational data does not necessarily become more abstract by mere transformation of frequency scales, or by transcription of Fø curves in terms of high and low points or contour shapes. The statement may seem trivial, but such a concept of "abstract" is not a matter of course, at least not in some of the present day phonological schools.

In this section I shall briefly present two different, current models of intonation, the hierarchial and the tonal sequence representation; discuss two particular claims about the shortcomings of the hierarchical theory; and discuss the present data in the light of both representations.

## A. THE HIERARCHICAL AND THE LINEAR INTONATION THEORIES

On a number of previous occasions, most recently in Thorsen (1983a), I have presented data and arguments in favour of a model of Standard Danish intonation in terms of a hierarchical, layered system of simultaneous, non-categorial intonational components of varying temporal scope, from sentences through prosodic phrases and prosodic stress groups to segments. I shall not repeat the reasoning here, but note that to the hierarchy of intonational components we may apparently add a textual contour. Similar views of the composition of fundamental frequency courses - with such modifications as language specific differences impose - are expressed about Japanese in Fujisaki et al. (1979), about French in Vaissière (1983), about Dutch in 't Hart and Collier (1979), and about Swedish in Gårding (1983) and Bruce (1977).[1] However, recently Bruce (1982) seems to be leaning more towards the basic concept of the linear theory of intonation as formulated by Pierrehumbert (1980).

Pierrehumbert (1980) analyses English intonation in terms of a linear sequence of categorially different, non-interacting pitch accents, which are associated with the stressed syllables of the utterance. They consist of either a high tone (H), a low tone (L), or bitonal combinations of the two. The inventory of pitch accents includes a (final) phrase accent and a boundary tone. The resulting description is very elaborate and very elegant, if occasionally somewhat complicated. Overall slope - if it is too steep to be accounted for by other factors - is handled by a downstep rule which lowers a H tone by a constant factor relative to the preceding H, in certain environments. Everything else being equal (such as the prominence relations between the stressed syllables of the utterance), this rule will create asymptotically declining slopes. Liberman and Pierrehumbert (1983) modify and expand the theory to include variations induced by changes in pitch range, as well as a lowering of the final pitch accent. Once the Fø value of the initial pitch accent is determined, such a theory requires no pre-planning, no look-ahead, on the part of the speaker for the execution of intonational phenomena, and "In general, we see no evidence that the Fo implementation for an entire phrase is necessarily laid out before speaking begins, even when the phrase is known in advance and fluently produced. All of our measurements can be modeled quite well on a left-to-right, plan-as-you-go basis." (Liberman and Pierrehumbert, 1983. The quotation is from the manuscript; I do not have the final version at the time of writing). This is also made explicit by Pierrehumbert (1983, p. 141):

"Superficially global trends arise from iterative application of these local rules." and "We also deny that intonation is built up in layers, by superposing local movements on a global component."

Pierrehumbert and Liberman see the data on Standard Danish intonation (specifically the fact that the slope of the declination varies with utterance type and function, cf. e.g. Thorsen 1978 and 1980c) as a challenge to the idea that intonation can be generated by local rules, i.e. rules with a narrow domain or temporal scope (extending backwards). They meet the challenge with a proposal for a downstep rule (whose specific formulation is different from the one they posit for English) where either the downstep factor or the reference line (which is part of the formula for target value calculation) is allowed to vary with sentence type. (Thorsen 1983a+b address this specific issue.)

The essential difference between the two types of intonation model, a hierarchical versus a linear one, is not merely formal or notational but conceptual or ideological, in a hackneyed term. One theory claims - or implies tacitly - that certain, gross aspects of the intonation of an utterance are anticipated and planned for, or laid out, at the moment the utterance is initiated. The other theory denies the existence of such preplanning and look-ahead, except that variation in total length may be staked out in the value of the very first pitch accent. One sees the fundamental frequency contours of utterances as a composite of contributions from, among others, an intonational and an accentual component. In the linear theory intonation is made up of, consists of, a sequence of pitch accents. These aspects, the layered versus simply sequential and the global versus local planning of tonal events, are what Pierrehumbert (1980) and Liberman and Pierrehumbert (1983) are most concerned with when they opt for the linear, tonal sequence model. They do not concern themselves so much with, e.g., descriptive adequacy. This and other problems are treated extensively by Ladd (1983a and b).

## B. LADD'S CRITICISM OF THE HIERARCHICAL REPRESENTATION

Ladd (1983a and b) modifies Pierrehumbert's (1980) theory while staying within its general framework of linearity and locality. He restricts the inventory of pitch accents and replaces the downstep rule with a downstep feature. Ladd adds a couple of other features and claims (I think rightly) thereby to achieve a less complex description of intonational phenomena. Ladd (1983a) specifically treats the generation of overall declining intonation contours without recourse to preplanned grid-lines (cf. Gårding, 1983), or baselines and plateau's (cf. Vaissière, 1983) or other such abstract contour shapes, among which he counts the sentence intonation component which I have posited for Standard Danish (e.g. Thorsen, 1978 and 1983a). In his discussion of the relative merits of the two types of intonation models, he makes a couple of statements about the hierarchical model which I would like to address, because they are likely to obscure the issue.

Ladd (1983a, p. 40) sees a difference in the two theories in their weighting of phonetic explicitness against functional relevance. Functional generalizations are conveniently expressed in descriptions of overall shape or slope of sentence contours in the hierarchical model, but phonetic detail is ironed out or lost, he says. Once such detail is admitted, Ladd claims, the number of different sentence intonation types multiplies rapidly, and he sees no basis for extracting further, explicit functional generalizations. Ladd refers here to the results on sentence intonation in long sentences in Danish (Thorsen 1983a): At and above five prosodic stress groups, sentence intonation contours are not straight lines but become bumpy, due to partial resetting of the declining slope. The theme is picked up again later (p. 50), where Ladd asserts that not all hierarchical models assume that the lines connecting accents should be straight, since "*...Thorsen..admits the existence of irregular intonation lines, though she is at a loss to account for them or fit them into her overall view of sentence intonation.*" This conclusion is Ladd's, not mine; moreover, it is incorrect. The fact that long intonation contours are bumpy does not exclude them from being globally preplanned, and it does not prohibit the extraction of explicit functional generalizations (about sentence type and the like). But the intonational realization of the same function, say terminal declarative, may come out in different shapes, or variants, depending - inter alia - on the length of the utterance. A complication arises from the fact that the location of the bumps (the resettings) in an intonation contour is determined in an intricate and - so far - nowhere near fully investigated interplay with the syntax and semantics of the utterance. This is (probably) what Ladd alludes to when he claims that "irregular" contours cannot be accounted for. However, the problems of prosodic, syntactic and semantic interplay are not specific to the hierarchical model of intonation, and they are problems that await a massive amount of empirical studies for their solution. Thus, I would still like to claim that the variation in sentence intonation contour shape is principled (allowing, of course, for inter as well as intra speaker variation or "noise"); i.e. it is context determined, bound variation, even though the governing principles are not all known to us yet.

## C. PRE-PLANNING TEXTUAL CONTOURS

Thorsen (1983a+b) contain a discussion of some specific drawbacks that I think the tonal sequence theory faces in the description of Danish intonation. Suffice it here to repeat the conclusion: If the downstep rule or feature is to capture the amount of variability and complexity in intonation contours in simple sentences of varying length and syntactic and semantic make-up, it must be a scalar and continuous feature which will often have to take different values within one and the same phrase or sentence. I.e. it must have at least some of the properties it presumably was intended to rid intonation analysis of. - I am of course fully aware that the hierarchical model faces just as serious problems where the formalization and description of fundamental frequency data are concerned, but I repeat that the real difference between the two theories is neither formal nor notational but lies in their assumption of hierarchical, globally preplanned versus sequential, locally determined tonal events.

Before I proceed, let me stress the fact that the texts accounted for in section II were composed solely of declarative sentences and clauses. I do not know how a series of questions, strung together into a text, would have behaved, or alternating declarative and interrogative sentences. Nor can I make any claims about longer texts or texts composed of shorter or longer constituent components. And last, but not least, the present analysis pertains to highly planned (read) speech. The production of an ensemble of utterances, constituting a coherent text, may be much less regular in free, spontaneous, non-monitored speech.

The discussion to follow can be limited to the behaviour of the lower lines, since the upper lines are closely correlated with and can be derived from them, cf. section III.B.1 above. In itself this fact may be taken as evidence for the layered composition of fundamental frequency: the lower line is the connection of the stressed syllables of an utterance, the intonation contour proper in my framework. The upper line connects the first post-tonic syllable in the prosodic stress groups. A stress group may or may not have post-tonics, i.e. there may or may not be centres of suspension for an upper line. However, the lower line is oblivious to this vacillation. The stressed syllables are frequency scaled in relation to each other, without regard to the presence or not of any "highs" in the surroundings. The lower line is a constant, ceteris paribus, upon which high-falling patterns are superposed whose range and extent and detailed shape are determined by the context (earlier or later, on a more or less sloping contour) and by the duration of the prosodic stress group (in terms of number of syllables). See further Thorsen (1980b and 1984).

First of all, I find it difficult to accept that patterns such as those in figure 6 could be obtained without a fair amount of pre-planning and look-ahead on the part of the speaker. More specifically, the different arrangement of the individual sloping (lower) lines according to the strength of the boundary between otherwise identical sentences in a text can only be the result of look-ahead and pre-planning, it seems to me. The step down between the first and second, and the second and third stressed syllable, i.e. the slope of the line connecting them, is different when the unit is bounded by a sentence boundary versus a clause boundary, in a manner which makes three successive coordinate clauses less differentiated and more coherent relative to the overall declination exhibited by the whole text than three terminal sentences in succession, cf. figure 6.

The other - as I see it - crucial point ties up directly with a very explicit hypothesis about the domain of phonetic implementation rules in Liberman and Pierrehumbert (1983 - once again the quotation is from their manuscript):

*"Now comes the conjectural part: we insist that the computation of any parameter or object Y[i] can only depend on the "accessible" properties of Y[i] and Y[i-1], where Y[i-1] means the immediately previous object of the same type (if any). Thus, pitch accent can look back to previous pitch accent, phrase to previous phrase, etc.*

*... A restriction of this type has a certain functional value, for both speakers and hearers; speakers can get on with the task at hand without knowing all the details of what follows, while hearers can in principle complete the phonetic processing of what they have heard up to any given point in the stream of speech.*

*Many apparent instances of anticipatory effects are known. Unless our conjecture is wrong, all such cases must turn out to be explained either by feature spreading at the phonological level, or else by computation of some parameter of a higher level constituent."*

I will not argue the functional value of this conjecture. But, functional or not, I think it is clearly contradicted by the following fact (cf. III.B.1. and B.2. above): the slope of a sentence or clause contour, occupying a given slot in the left-to-right order of components, takes different values according as it is succeeded by one or more components, or not. Thus, an isolated sentence has a steeper slope, the downstep is greater between successive stressed syllables, than when the same sentence is succeeded by one or more (that is immaterial) sentences in the same text. And sentence or clause no. 2 from the left is steeper when it is simultaneously text final than when it is text medial. In fact, at least three positions are distinguished in longer texts: initial, medial, and final. Note specifically that the variation with text length (isolated sentence versus multi and two versus three sentence text) is not primarily expressed in a lower or higher beginning of the whole contour. (Such a phenomenon could be accommodated in Liberman and Pierrehumbert's representation by manipulating the overall range and the derived reference line against which the initial pitch accent target value is computed.) On the contrary, beginning and end points are essentially constant across texts of different length. The overall textual contour slope varies, and so do the slopes of its individual components, according not only to preceding, but also to upcoming events.

This last observation differs from Bruce's (1982) conclusion. As mentioned in sectionIII.E above he found that an Fø minimum in the first accent after a text medial sentence boundary continues at the same value that the corresponding Fø minimum had in the last accent before the boundary. He says (p. 285): *"The observed tonal adaptation of the earlier part of a succeeding sentence to the later part of a preceding one within the same textual unit can be taken as an indication that there is a partial rather than a total preplanning of the global Fø course in relation to the length of the actual text unit."* If - as I speculated above - this particular intonational configuration of successive sentence components is an artifact of the material, and if it would not be replicated with longer individual sentence components, then Bruce's hypothesis of partial rather than total preplanning is no longer so easily sustained.

Similar contradictory evidence against exclusively backwards extending tentacles was found in intonation contours in declarative terminal sentences of varying length (Thorsen 1983a). The second stressed syllable is lower, relative to the first stressed syllable, when it is simultaneously sentence final than when another stressed syllable follows, and it is lower succeeded by one than by two stressed syllables; likewise for the third stressed syllable. The phonetic implementation is demonstrably not independent of or insensitive to succeeding events.

If data of the kind presented in this paper is to be accommodated within the theory of locally determined tonal sequences, say Liberman and Pierrehumbert in the last paragraph of the quotation above, it must be explained either as feature spreading at the phonological level, or else by computation of some parameter of a higher level constituent. I cannot see that feature spreading at the phonological level is anything but the linguist's way of formally accounting for the preplanning of articulatory events; it does not "explain" these events. And I cannot see that referring anticipatory effects to a higher level constituent is anything but admitting the existence of hierarchically organized tonal events.

Experiments like the present one are the most favourable we can imagine if we want to know what speakers can achieve by way of precision and constancy in speech production tasks. Not only is the speech monitored and planned for them, but they get to repeat the same items over and over (though not in straight succession). Even so, the signals we pick up are "noisy". This intra speaker variation (as well as the more considerable variation among speakers) is perhaps less disturbing, less alien, to the representation of intonation in terms of gross, global events which are filled in with more detailed contributions (i.e. of lesser temporal scope). Precisely because the actual execution of tonal events is the result of interacting larger and smaller scale components (because they are expressed in the same medium), we should not expect these events to be performed with mathematical precision. It seems to me that the tonal sequence theory, and the concept that gave birth to it, focus too heavily on just such a mathematical precision and thus ignores or at least underrates the plasticity involved in all aspects of speech production. With this last paragraph I am also implying that free, non-monitored texts will hardly be as regular as the ones investigated here. That is not to say that it would not be an interesting and worthy object of study - quite the contrary.

## V. NOTE

1. I am aware that Ladd (1983a and b) sees 't Hart and Collier's description of Dutch as belonging within the tonal sequence theory of intonation, and Gårding's and Bruce's Swedish model as a compromise model between the hierarchical and linear theories. I disagree with Ladd's interpretation. This controversy is, however, not central to my present purpose, and will not be dealt with any further here.

ACKNOWLEDGEMENT

My thanks are due to Jørgen Rischel for valuable comments. Jeanette Holtse competently performed the measuring and graphic work.

REFERENCES

Bruce, G. 1977: "Swedish word accents in sentence perspective", *Travaux de l'Institut de Linguistique de Lund 12*, p. 1-155

Bruce, G. 1982: "Textual aspects of prosody in Swedish", *Phonetica 39*, p. 274-287

Cooper, W.E. and Sorensen, J.M. 1981: *Fundamental Frequency in Sentence Production*, (New York: Springer Verlag)

Fujisaki, H., Hirose, K. and Ohta, K. 1979: "Acoustic features of the fundamental frequency contours of declarative sentences in Japanese", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 13*, p. 163-173

Gårding, E. 1983: "A generative model for intonation", in: A. Cutler and D.R. Ladd (eds.) *Prosody - Models and Measurements*, (Berlin: Springer-Verlag), p. 11-25

't Hart, J. and Collier, R. 1979: "On the interaction of accentuation and intonation in Dutch", in: E. Fischer-Jørgensen, J. Rischel and N. Thorsen (eds.) *Proc. 9th Int. Cong. Phonetic Sciences, vol. II*, (Copenhagen: Institute of Phonetics, Copenhagen University), p. 395-402

Hombert, J.-M. 1977: "Development of tones from vowel height?", *J. Phonetics 5*, p. 9-16

Jeel, V. 1975: "An investigation of the fundamental frequency of vowels after various consonants, in particular stop consonants", *Ann. Rep. Inst. Phon. Univ. Copenhagen 9*, p. 191-211

Ladd, D.R. 1983a: "Peak features and overall slope", in: A. Cutler and D.R. Ladd (eds.) *Prosody - Models and Measurements*, (Berlin: Springer-Verlag), p. 39-52

Ladd, D.R. 1983b: "Phonological features of intonational peaks", *Language 59*, p. 721-759

Lehiste, I. 1975: "The phonetic structure of paragraphs", in: A. Cohen and S.G. Nooteboom (eds.) *Structure and Process in Speech Perception*, (Berlin: Springer-Verlag), p. 195-206

Lehiste, I. 1979: "Perception of sentence and paragraph boundaries", in: B. Lindblom and S. Öhman (eds.) *Frontiers of Speech Communication Research*, (London: Academic Press), p. 191-201

Lehiste, I. 1980: "Phonetic characteristics of discourse", *Ac. Soc. Japan, Transactions of the committee on speech research, April 1980*, p. 25-38

Lehiste, I. and Wang, W. S.-Y. 1977: "Perception of sentence boundaries with and without semantic information", in: W.U. Dressler and O.E. Pfeiffer (eds.) *Phonologica 1976, Innsbrucker Beiträge zur Sprachwissenschaft* (Institut für Sprachwissenschaft der Universität Innsbruck), p. 277-283

Liberman, A.M., Cooper, F.S., Shankweiler, D.S., and Studdert-Kennedy, M. 1967: "Perception of the speech code", *Psychological Review 74*, p. 431-461

Liberman, M. and Pierrehumbert, J. 1983: "Intonational invariance under changes in pitch range and length", in: M. Aronoff and R. Oehrle (eds.) *Language Sound Structure*, (Cambridge: MIT Press)

Nakatani, L. 1975: "Sentence prosody in isolated and in story contexts", *J. Acoust. Soc. Am., Supplement 1*, p. 120

Petersen, N. Reinholt 1974: "The influence of vowel height on the perception of vowel duration in Danish", *Ann. Rep. Inst. Phon. Univ. Copenhagen 8*, p. 1-10

Pierrehumbert, J.B. 1980: *The Phonology and Phonetics of English Intonation*, M.I.T. Doctoral dissertation

Pierrehumbert, J.B. 1983: "Linguistic units for Fo synthesis", *Abstracts of the 10th Int. Cong. Phonetic Sciences*, (Dordrecht: Foris Publications), p. 137-144

Thorsen, N. 1978: "An acoustical investigation of Danish intonation", *J. Phonetics 6*, p. 151-175

Thorsen, N. 1979: "Interpreting raw fundamental frequency tracings of Danish", *Phonetica 36*, p. 57-78

Thorsen, N. 1980a: "Word boundaries and Fo patterns in Advanced Standard Copenhagen Danish", *Phonetica 37*, p. 121-133

Thorsen, N. 1980b: "Neutral stress, emphatic stress, and sentence intonation in Advanced Standard Copenhagen Danish", *Ann. Rep. Inst. Phon. Univ. Copenhagen 14*, p. 121-205

Thorsen, N. 1980c: "A study of the perception of sentence intonation - Evidence from Danish", *J. Acoust. Soc. Am. 67*, p. 1014-1030

Thorsen, N. 1983a: "Standard Danish sentence intonation - Phonetic data and their representation", *Folia Linguistica 17*, p. 187-220

Thorsen, N. 1983b: "Two issues in the prosody of Standard Danish", in: A. Cutler and D.R. Ladd (eds.) *Prosody - Models and Measurements*, (Berlin: Springer-Verlag), p. 27-38

Thorsen, N. 1984: "Variability and invariance in Danish stress group patterns", *Ann. Rep. Inst. Phon. Univ. Copenhagen 18*, p. 13-43. To appear in *Phonetica 41*

Uyeno, T., Hayashibe, H. and Imai, K. 1979: "On pitch contours of declarative, complex sentences in Japanese", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 13*, p. 175-187

Uyeno, T., Hayashibe, H. and Imai, K. 1981: "Syntactic structures and prosody in Japanese: A study on pitch contours and the pauses at phrase boundaries", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 15*, p. 91-108

Vaissière, J. 1983: "Language-independent prosodic features", in: A. Cutler and D.R. Ladd (eds.) *Prosody - Models and Measurements*, (Berlin: Springer-Verlag), p. 53-66.