(ARIPUC 12, 1978)

ON THE IDENTIFICATION OF SELECTED DANISH INTONATION CONTOURS

Nina G. Thorsen

Abstract: 10 Copenhagen and 4 non-Copenhagen speakers identified 15 human utterances, differing only in their fundamental frequency course, as being either declarative, non-final, or interrogative (forced choice). Responses are very clearly correlated with Fo: the most steeply falling intonation contours are identified as being declarative, the least falling (i.e. "flat") ones as being interrogative, and contours in the middle of the continuum as being non-final. Copenhagen speakers clearly perceive three categories, whereas non-Copenhagen speakers seemingly operate with only two. Several, mutually interdependent, parameters in the Fo course may account for the results, the two most powerful ones, however, being the levels of the last stressed and the succeeding unstressed syllable in the utterance. In a subsequent experiment, 7 Copenhagen speakers identified the same utterances as being either declarative or non-declarative. The majority of the (formerly) non-final sentences were now labelled non-declarative, rather than being split into partly declarative, partly non-declarative categories.
When a subset of the same utterances were mutilated, identification criteria changed, and identification deteriorated almost progressively with the number of syllables being cut away from the end of the utterance (but not seriously - until only the first stress group remained), whereas syllables cut away from the beginning hardly affected identification at all.

1. Introduction

In a previous volume of ARIPUC the results of a preliminary analysis of intonation in Advanced Standard Copenhagen Danish (ASC) were presented. They were summarized in a model for fundamental frequency in short sentences, cf. fig. 1. For a detailed account of the procedure that led to the formulation of this model, the reader is referred to Thorsen (1976). However, those features that are relevant to the present experiment will be

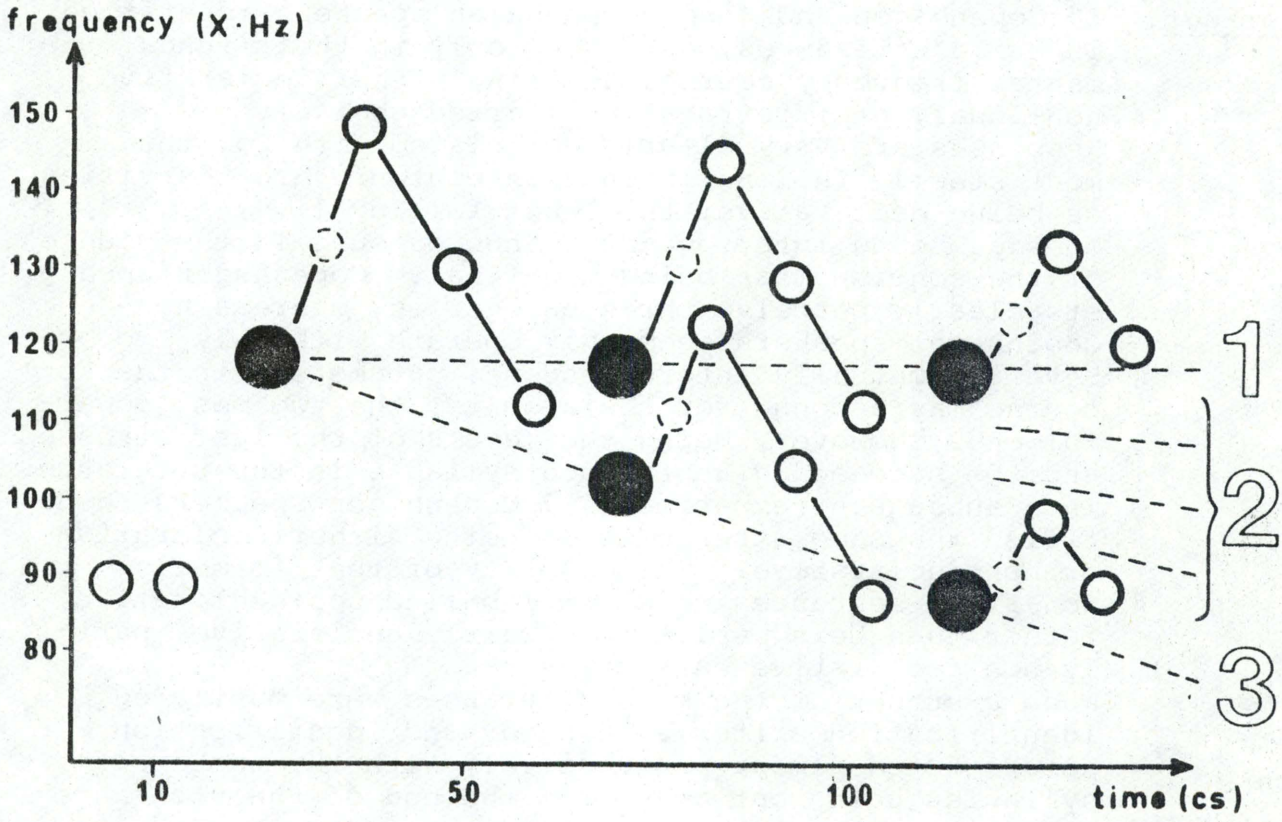


Figure 1

A model for fundamental frequency in short sentences in Advanced Standard Copenhagen Danish. 1: statement questions, 2: interrogative sentences with word order inversion and/or with interrogative particle, and non-final periods (variable), 3: declarative sentences.
The heavy circles represent stressed syllables, the empty circles unstressed syllables, and the broken circles represent syllables with assimilated or elided /ə/.
The full lines represent the fundamental frequency pattern associated with stress groups and the broken lines denote the intonation contour.

briefly outlined here: An underlying assumption is that the complex course of Fo in an utterance is the outcome of a superposition of several components.

(1) A sentence component which supplies the INTONATION CONTOUR.
(2) The contour is overlaid by a stress group component which furnishes the STRESS GROUP PATTERNS (both exemplified in fig. 1).
(3) To the resultant of those two components is added a stød component, rendering STØD MOVEMENTS. (However, as stød words had been excluded from the material, the model does not include this particular feature.) These first three components are language specific and thus "speaker controlled". (4) Finally, intrinsic fundamental frequencies of segments and coarticulatory variations at segment boundaries supply a MICROPROSODIC COMPONENT, which is not consciously controlled by the speaker, but due to inherent properties of the speech production apparatus. - The same point of view about "layers" in intonation has been expressed previously by several authors, see e.g. Bolinger (1970), Bruce (1977), Cohen and t' Hart (1967), Collier and t' Hart (1975), t'Hart (1966), t'Hart and Cohen (1973), and Lehiste and Peterson (1961), and for a more thorough account, see Thorsen (forthcoming).

As may be seen from the full lines of fig. 1, the stress group pattern can be described as a (relatively) low stressed syllable followed by a high-falling tail of unstressed syllables. This pattern is a predictable and recurrent entity, allowing, however, for contextual variations in the magnitude of the rise from stressed to unstressed syllable, which decreases from the beginning to the end of the utterance. In fact, it was this observation which led to the definition of the stress group in Danish as a stressed syllable plus all succeeding unstressed syllables (within the same non-compound sentence), irrespective of intervening word- or morpheme boundaries, and it also led to the definition of the intonation contour as the course described by the stressed syllables alone, cf. the dotted lines of fig. 1. The same concept of the intonation contour can be found in Bolinger's (1958, 1970) treatment of American English, and it is similar to the "declination" line from which the "hat" patterns set off in Dutch, cf. e.g. Collier and t'Hart (1975) and t'Hart and Cohen (1973).

Fig. 1 may be accounted for in a different manner, namely in terms of top-lines and base-lines which are tangents to the maxima and minima, respectively, in the Fo course. This would be in accordance with the way Bruce and Gårding (1978) describe Swedish, and, likewise, with the description by Breckenridge and Liberman (1978) of American English, except that in Swedish and American the top-line is tangent to the stressed syllables, and the base-line to the unstressed syllables. A feature common to all three languages is the fact that both lines decline (more or less steeply), but the top-line declines faster than the base-line, so that top- and base-line together create a wedge-shape. (Assuming that the stressed syllables determine the intonation contour, this would mean that the top-line carries the perceptual cue in Swedish and American, where it is the base-line in ASC Danish.) One objection that could be raised towards applying this description to Danish is that the stressed syllables do not always constitute the minima in the Fo course, - sometimes the "base-line" (i.e. the intonation contour) is transgressed by the tail of unstressed syllables, see e.g. contour "1" of fig. 1.

The intonation contour tends to vary systematically with sentence type, as suggested by fig. 1: declarative sentences having the most steeply falling contours (about 25%/sec), at one extreme, and statement questions (i.e. questions where only the intonation contour signals their interrogative function) having "flat" contours, at the other extreme. In between these two are found other types of questions as well as non-final periods. Further, there seems to be a certain trade-off between syntax[1] and intonation contour: the more syntactic information is contained in the sentence about its non-final or interrogative function, the more declarative-like, i.e. the more steeply falling, is its intonation contour, cf. Thorsen (1976, pp. 134-135). Such a trade-off has also been observed for other languages, see e.g. Bolinger (1964), Cohen and t'Hart (1967), Daneš (1960), von Essen (1956), Hadding-Koch (1961), and Mikoš (1976). - It was this observation that led to the formulation of the following three questions, which the experiments to be reported below were designed to answer:

---

1) I do not wish, by this use of the terminology, to exclude the possibility of treating intonation as an integral part of the syntax of the language, and 'syntax' (and 'syntactic') should therefore just be regarded as a convenient abbreviation for "other signals, such as word order inversion, interrogative particles, and the like".

(1) will listeners identify three types of contours, in correlation with the actual course of Fo?
(2) if the answer to the first question is affirmative: are these categories linguistic, rather than purely phonetic?
(3) how early, or how late, in the sentence are the various contours perceptually differentiated?

2. Test 1

Test 1 attempts to answer the first question above, whether listeners will identify three categories of contours. The search for two or more (linguistic and/or attitudinal) categories of sentence intonation is anything but new, see e.g. Daneš (1960), Delattre et al. (1965), von Essen (1956), Gårding and Abramson (1965), Hadding-Koch and Studdert-Kennedy (1963, 1964, 1965, 1974), Isačenko (1965), Isačenko and Schädlich (1963), Johansson (1978), Mikoš (1976), Studdert-Kennedy and Hadding (1972, 1973), and Uldall (1960, 1961). The primary motivation for conducting yet another experiment on the identification of intonation contours is the fact that it has not been done with Danish material before. Furthermore, the material and procedure deviate somewhat from that of previous investigations.

2.1 Test material - test 1

A small part of the material recorded for the original analysis was selected for identification. I.e. natural speech was employed. The reason for this choice lies first and foremost in a curiosity to see how the contours, as actually produced by a speaker, would be identified, and whether the obviously and rather systematically different contours can serve a perceptual and linguistic purpose. Secondly, experiments with real speech would be a natural prerequisite to ones with synthetic speech, i.e. the relevant parameters for systematic variation in synthetic speech would appear in this fashion. Thirdly, the objection that could, rather maliciously and not wholly reasonably perhaps, be raised against the many investigations conducted with synthetic speech (or "semi-synthetic" speech, i.e. vocoder reproduced segments with externally controlled Fo) that listeners are exposed to stimuli that could never occur in real speech, is muted from the outset. (Note that all of the perceptual experiments cited in the bibliography, except for parts of the material of Gårding and Abramson (1965) and parts of Johansson's (1978) stimuli, have been performed with synthetically produced stimuli.)

A further advantage is the fact that the stimuli presented to subjects here are fairly long, comprising three stress groups (7 syllables in all), whereas the studies by Hadding-Koch and Studdert-Kennedy (1963, 1964, 1965, 1974) and Studdert-Kennedy and Hadding (1972, 1973), which most closely resemble, in outline and purpose, the present study were conducted with utterances containing only one stressed syllable ('For Jane' and 'November', respectively).

2.1.1 Stimuli - test 1

A subset of the material for analysis of intonation contours contained a statement, 5 different types of questions, and 3 different types of non-final periods (i.e. 9 sentences in all, which had been recorded 5 times by four subjects), all variations on the same theme:

"..mange busser fra Tiflis .."

('.. many buses out of Tiflis ..'), cf. Thorsen (1976, pp. 91-92). The statement, one type of question, and one type of non-final period were selected from the recordings of one of the male subjects (SH), according to the following criteria: the averages over the five recordings of each type should be well spaced on the frequency scale, whereas the five recordings of each sentence type must show a certain dispersion, so as to create an at least quasi-continuous series of stimuli, from the most steeply falling to flat contours. (But this is, of course, where natural speech is at a disadvantage, compared to synthetic speech: it was not possible to procure a series of stimuli spaced completely equidistantly on the "slope-continuum".) These demands are best fulfilled by SH's non-terminal main clause:

"Der går mange busser fra Tiflis, så vi kan godt lade bilen stå."

('There are many buses out of Tiflis, so we may well leave the car.') and by the question with word order inversion:

"Går der mange busser fra Tiflis?"

('Are there many buses out of Tiflis?'). The statement was:

"Der går mange busser fra Tiflis."

('There are many buses out of Tiflis.'). The important point to note is that these sentences are identical from the [m] in "mange" to the [s] in "Tiflis", so the only difference across this stretch of 7 syllables is the intonation contour.

By means of a segmentator the sequence

mange busser fra Tiflis ([ˈmaŋ̍ ˈb̥usʌ fʁ̥ɑ ˈd̥ˢifl̥is])

was isolated from the 15 items. Only 50-70 ms of the final [s] was included, in order that no trace of the word "så" following the non-final period be detected. These stimuli are termed "D" (declarative), "NF" (non-final), and "I" (interrogative) and numbered from 1 to 5 (first, second, etc. recording) in the following. Fo tracings of the 15 stimuli are shown in fig. 2. (These tracings are not completely raw but have been processed so as to remove influence from surrounding obstruents, cf. Thorsen (1977,forthc.), but corrections for intrinsic Fo level differences have not been performed.) - In table 1 are given measures of the duration (in cs) of the segments and of each stress group. Differences between average durations of a particular segment or stress group across the three sentence types are small and in no case statistically significant.

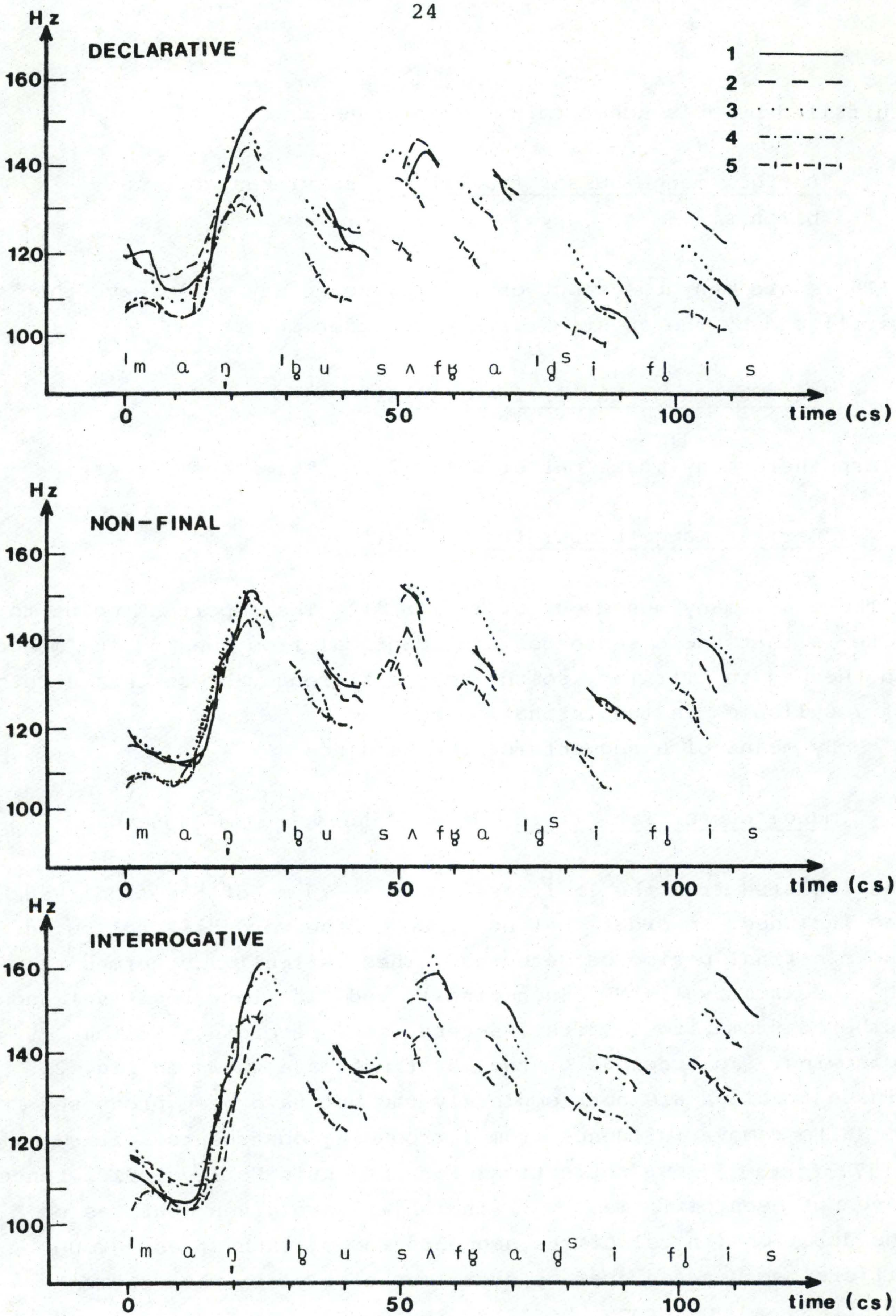


Figure 2

Fundamental frequency tracings
of the 15 stimuli of test 1.

Table 1

Durations (in cs) of segments and stress groups in the 15 stimuli.

	m	ɑ	ŋ(ə)	ˈmɑŋ(ə)	b̥	u	s	ʌ	fɣ̊	ɑ	ˈb̥usʌfɣ̊ɑ	d̥ˢ	i	fl̥	i	s	ˈd̥ˢifl̥is	Total
D-1	6	12	10	28	8	11	5	6	8	6	44	12	9	11	8	6	46	118
D-2	6			27	7	10	7	6	8	5	43	14	7	10	9	6	46	116
D-3	5	9	11	25	7	11	6	5	5	6	40	13	11	9	10	6	49	114
D-4	5	10	10	25	6	11	6	6	8	6	43	11	11	8	10	5	45	113
D-5	5	8	11	24	8	9	7	5	6	5	40	13	9	11	9	6	48	113
Average	5.4			25.8	7.2	10.4	6.2	5.6	7.0	5.6	42.0	12.6	9.4	9.8	9.2	5.8	46.8	114.8
NF-1	5			24	9	10	6	5	8	5	43	15	8	11	8	7	49	116
NF-2	5			26	9	8	6	5	8	5	41	15	7	13	5	7	47	114
NF-3	5	8	12	25	8	10	6	6	8	5	43	14	7	14	6	5	46	114
NF-4	5			25	8	9	5	6	9	4	41	13	8	10	9	6	46	112
NF-5	6	8	11	25	6	9	6	6	8	7	42	12	6	13	7	7	45	112
Average	5.2			25.0	8.0	9.2	5.8	5.6	8.2	5.2	42.0	13.8	7.2	12.2	7.0	6.4	46.6	113.6
I-1	6			27	10	10	6	5	10	5	46	13	8	11	10	6	48	121
I-2	5			27	9	10	4	6	8	5	42	13	9	10	10	5	47	116
I-3	5	11	12	28	8	10	5	8	9	5	45	10	10	9	8	6	43	116
I-4	5	9	13	27	7	8	8	7	8	4	42	12	8	11	9	5	45	114
I-5	5	8	11	24	7	10	5	7	9	4	42	12	10	11	9	5	47	113
Average	5.2			26.6	8.2	9.6	5.6	6.6	8.8	4.6	43.4	12.0	9.0	10.4	9.2	5.4	46.0	116.0

2.2 Test procedure - test 1

A test tape containing 5 different randomizations of the 15 stimuli, led by 5 dummies, i.e. 80 stimuli in all, was prepared. Stimuli were introduced by numbers (read by the author) with a 3 sec interval for responding. The total duration of one run of the test tape was 11 minutes.

Subjects were instructed that they would hear sequences cut out from a larger context (they were given several examples of possible contexts for each sentence type), and their task was to decide whether the cut originated from one of three sentence types: "declarative", "non-final", or "interrogative" (forced choice). - Three categories to choose from were, of course, motivated by the fact that the 15 stimuli represented 3 different types of sentences. Subjects were given forced choice with no possibility of responding "do not know" or of placing stimuli on a multi-valued scale, - mainly to facilitate the subsequent interpretation of results. - The test was presented twice, with at least one day's interval, giving a total of 10 responses to each stimulus by each subject. 10 subjects listened over head-phones, 4 subjects over loudspeakers in a class-room.

2.3 Subjects - test 1

14 subjects took the first test. 10 were trained phoneticians (among them the speaker), 3 were language students, with no phonetic training, and one was a trained singer with no background in phonetics. Two of the phoneticians and two of the language students were born and went to school outside the Copenhagen area, - the rest were genuine speakers of ASC. It turned out that Copenhagen and non-Copenhagen speakers behaved significantly different, and the two groups are treated separately in the following account of the results.

2.4 Results - test 1

First of all, there is no "order effect" to be detected with any of the subjects. That is, neither "correct" nor "incorrect" responses can be seen to be due to a particular stimulus/response immediately preceding. (An order effect would have been surprising, since stimuli occurred in 5 different randomizations on the tape, and since, further, a stimulus number was announced, by a different voice, and always with a declarative intonation, before each stimulus.) Secondly, the 10 Copenhagen speakers did not differ significantly among themselves, and they are, accordingly, pooled in figures and tables to follow.

Figs. 3 and 4 and tables 2 and 3 present the results, for Copenhagen and non-Copenhagen speakers, respectively. In the figures, stimuli have been reorganized according to the number of responses they received, from maximally interrogative, through maximally non-final, to maximally declarative.

It is immediately obvious from the shape of fig. 3 that there are stimuli in each of the three categories that have been well identified by the Copenhagen speakers: two declarative sentences (D-4 and D-5), two non-final periods (NF-1 and NF-2) and three interrogative sentences (I-1, I-2, and I-3) received 85% or more responses for the respective categories. It is equally obvious that non-Copenhagen speakers do not fare so well on the non-final category, which can hardly be said to have any peak in the identification function at all. Further, non-Copenhagen responses are, on the whole, far more overlapping: 9 stimuli received responses of all three kinds, as opposed to the Copenhagen speakers, who only gave all three kinds of responses to two stimuli (NF-3 and NF-2).

Stimuli that are grouped within parentheses in the two figures are stimuli whose distributions of responses in χ^2 tests turned out not to be significantly different. In both figures, 6 such groups appear, but the members of each group are not exactly identical. Neither is, by the way, the rank order of the stimuli, i.e. NF-3 is shifted one to the left and NF-4 is shifted two to the left in fig. 4 as compared to fig. 3.

The non-Copenhagen speakers thus seem to have been in rather great doubt about the second category, the non-final one, and one might cautiously conclude that these speakers simply do not operate with more than two categories. (However, 4 speakers, who were not even representatives of the same non-Copenhagen dialect,

Table 2

10 Copenhagen speakers' response to the 15 stimuli in test 1.

response / stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	84	16		100
D-2	75	25		100
D-3	79	21		100
D-4	97	3		100
D-5	100			100
NF-1		88	12	100
NF-2	4	90	6	100
NF-3	2	79	19	100
NF-4	81	19		100
NF-5	34	66		100
I-1		5	95	100
I-2		11	89	100
I-3		15	85	100
I-4		61	39	100
I-5		57	43	100
Total	556	556	388	1500

Table 3

4 non-Copenhagenen speakers' response to the 15 stimuli in test 1.

response / stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	23	16	1	40
D-2	18	20	2	40
D-3	21	17	2	40
D-4	31	8	1	40
D-5	39	1		40
NF-1		27	13	40
NF-2	7	23	10	40
NF-3	2	15	23	40
NF-4	16	23	1	40
NF-5	12	25	3	40
I-1			40	40
I-2			40	40
I-3			40	40
I-4		20	20	40
I-5	2	15	23	40
Total	171	210	219	600

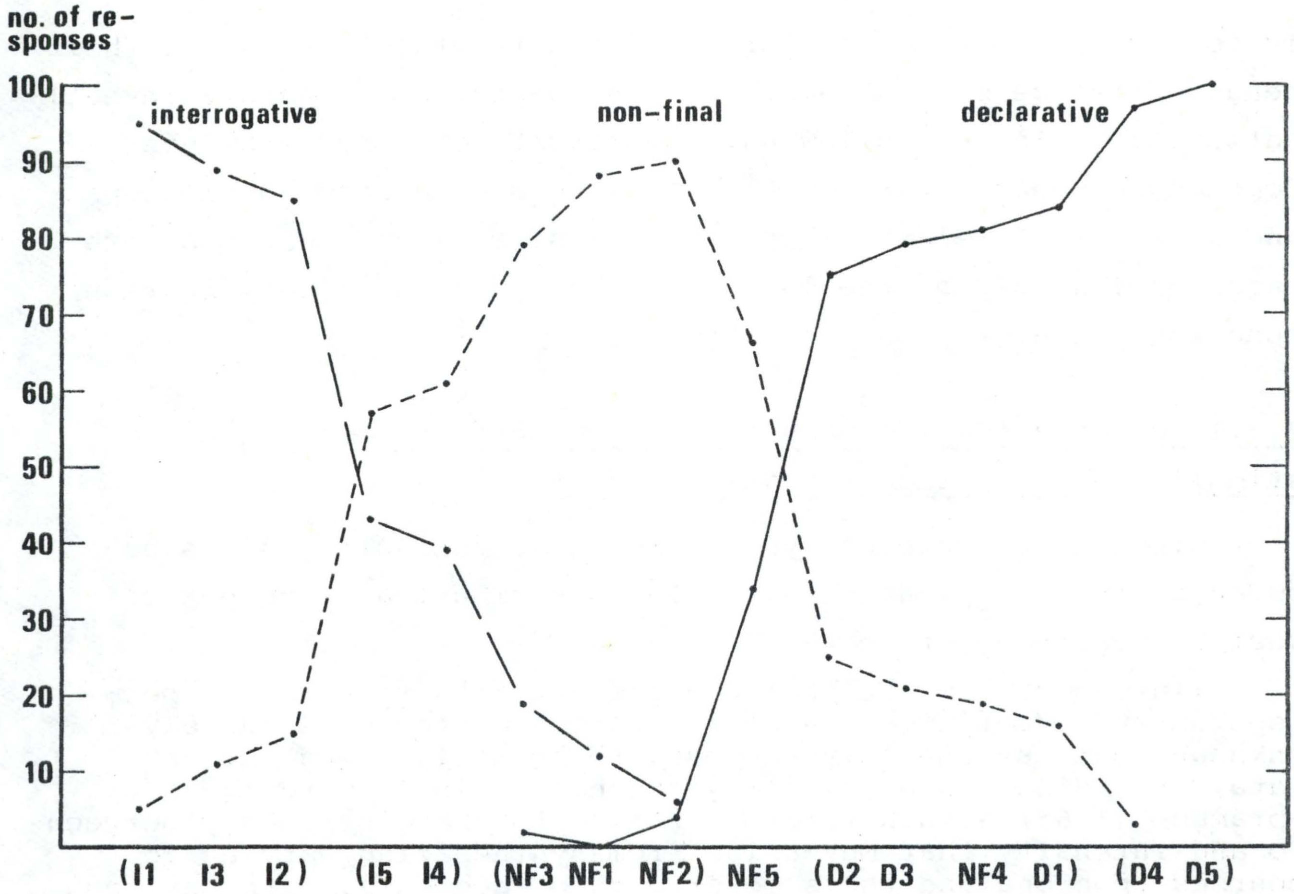


Figure 3

Identification functions for the 15 stimuli in test 1, as identified by 10 Copenhagen speakers.

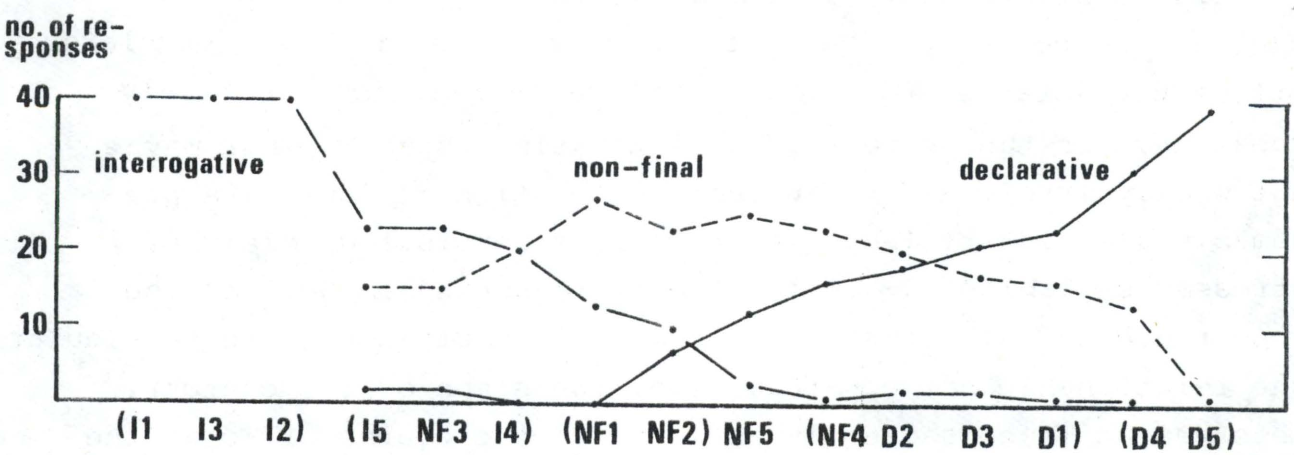


Figure 4

Identification functions for the 15 stimuli in test 1, as identified by 4 non-Copenhagen speakers.

is too small a population for any definite statements about their behaviour to be made - and this point warrants a separate investigation.) - In the following, the results obtained from non-Copenhagen speakers are omitted from consideration: conclusions and analyses of relationships between stimuli and responses are based exclusively on the 10 Copenhagen speakers' identification functions, i.e. fig. 3 and table 2.

2.4.1 Correlations between physical properties of the stimuli and their identification - test 1

Since there were no systematic durational differences between stimuli, explanations for the distribution of responses must be searched for in the Fo course.

(Intensity has not been considered. Its role in the perception and identification of intonation contours is largely unknown, but is generally supposed to be minimal and, at any rate, subordinate and ancillary to that of Fo. - Gårding and Abramson (1965) attach some importance to the discrepancy between Fo and intensity that may arise from synthesizing varying Fo courses upon one and the same stretch of segments, with one and the same distribution of intensity, and Breckenridge and Liberman (1978) found that even small amplitude adjustments have rather large effects upon judgments of the pitch of the second stressed vowel in a "maMAmamaMAma" sequence. Both studies, however, employed synthetic speech and there is as yet no reason to doubt that in natural speech fundamental frequency is the dominant factor in the perception of intonation.)

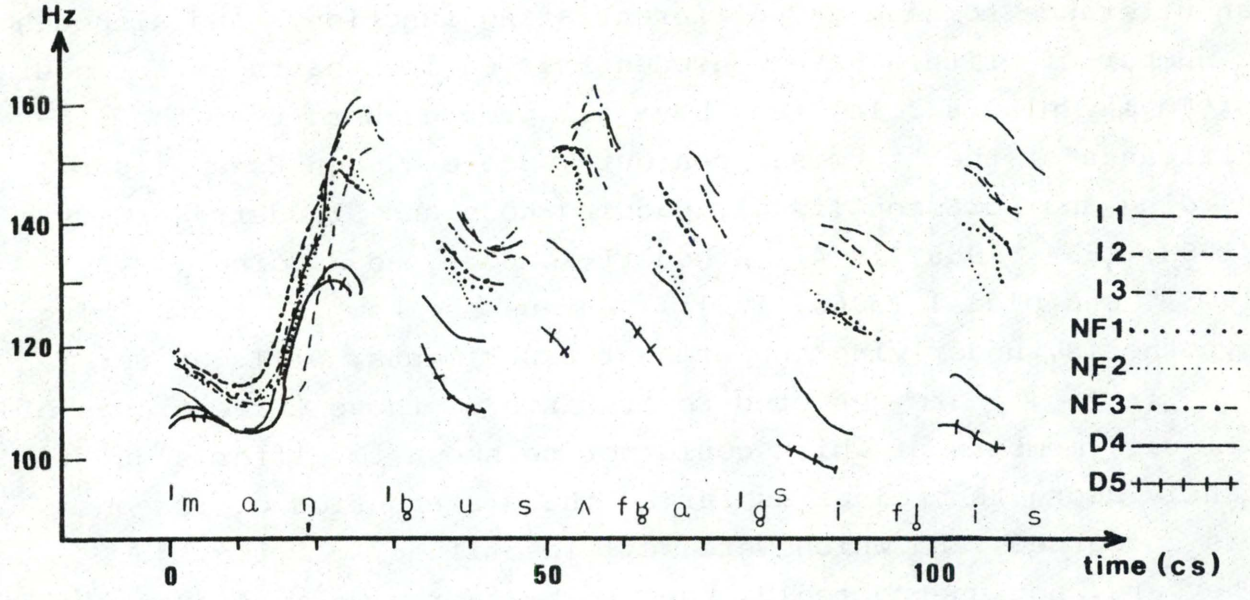
Since stress group patterns are recurrent entities, cf. fig. 1, the course of the unstressed syllables is to a very large extent predictable and might therefore be redundant, strictly speaking, for the perception of intonation contours (but maybe not wholly irrelevant). Further, intonation contours are presumably also identifiable in utterances consisting solely of stressed syllables. - This line of reasoning emphasizes the importance of the stressed syllables in an utterance, in particular the relationship between them, i.e. the slope of the contour. Less weight is consequently assigned to the course of Fo at the very end. This is in explicit contradistinction to the viewpoint expressed by Daneš (1960), Isačenko (1965), and Isačenko and Schädlich (1963) who considered only the terminal Fo course in an utterance to have any differentiating function. And although a number of writers have conceded that earlier parts of the course of Fo may play a role, they have all proceeded to test the significance of the "terminal contour", see e.g. von Essen (1956), Gårding and Abramson (1965), Hadding-Koch and Studdert-Kennedy (1963, 1964, 1965, 1974), Studdert-Kennedy and Hadding (1972, 1973), and Uldall (1960, 1961). Let us see how far the working hypothesis, underlying the present experiments, will get us:

In fig. 5 are depicted Fo tracings of those three groups of stimuli, members of which could not be shown to differ significantly among themselves as far as the distribution of responses was concerned, and which were well identified, cf. fig. 3. They are, not unexpectedly, seen to be fairly well separated, and this separation is most evident in the final stressed vowel ("Tiflis"), less evident in the second stressed vowel ("busser") but still without any overlap between the three groups. The unstressed syllables of the second and third stress groups tend to distribute themselves in the same manner as the stressed syllables, but some overlap is to be found. It is hard to find any order at all in the first stress group. This is all as it should be, granted that different contours, departing from the same point, are more clearly separated, the further we progress in time, cf. fig. 1.

In fig. 6 are shown Fo tracings of stimuli that have received ambiguous responses (i.e. stimuli at or near the cross-over points of the identification functions of fig. 3: I-4, I-5 and NF-5, D-2) as well as one well identified stimulus from each category (I-1, NF-1, D-5). I-4 and I-5 are well removed from I-1, and much closer to NF-1 (especially in the third stress group). Likewise, NF-5 and D-2 are well removed from D-5 (but not particularly close to NF-1).

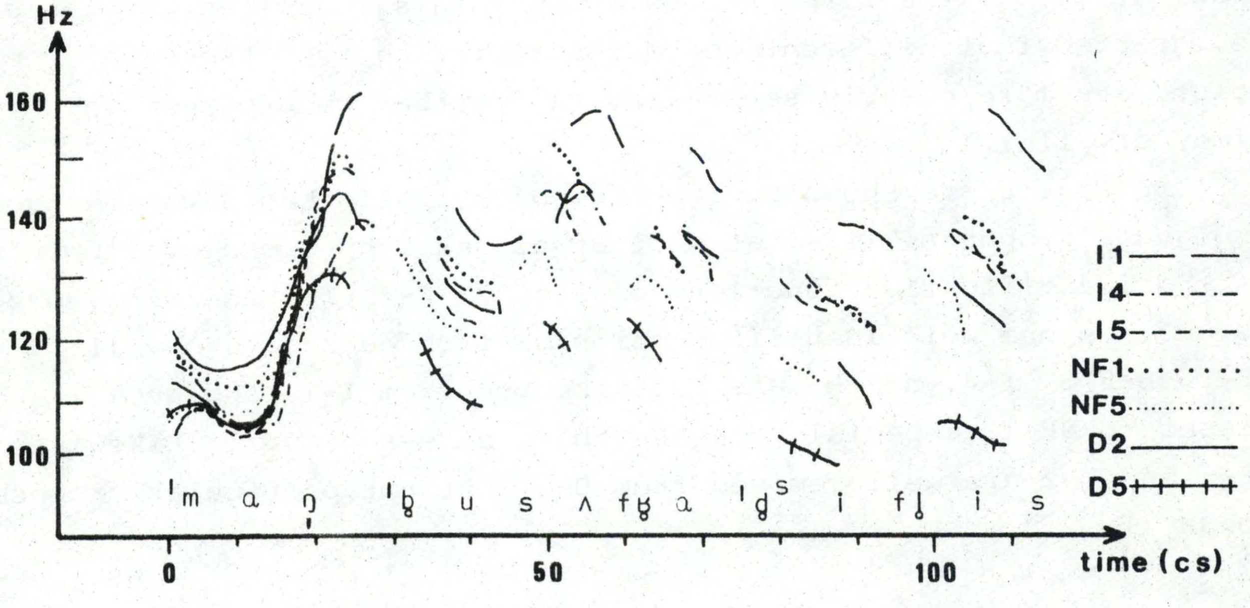
2.4.1.1 A closer look at stimuli

Figs. 5 and 6 may thus account for the gross trends in the results, but they cannot account for the finer distinctions we observe in the identification functions. In order to find one



<u>Figure 5</u>

Fundamental frequency tracings of groups of stimuli that have been well identified by 10 Copenhagen speakers, cf. fig. 3.



<u>Figure 6</u>

Fundamental frequency tracings of stimuli at or near the cross-over points of the identification functions of fig. 3 (I-4, I-5 and NF-5, D-2) as well as stimuli that have been well identified (I-1, NF-1, D-5) by 10 Copenhagen speakers.

or more relevant parameters, i.e. parameters which will yield a high correlation with the distribution of responses, table 4 was prepared. It contains a number of measures of various points in the Fo tracings, as well as certain relations which have been calculated on the basis of these measures.

Measuring Fo is not a wholly uncontroversial business. As a rule, in all the vowels, except [ˈɑ], that point has been measured which lies at 2/3 of the distance in time from the beginning of the vowel, in accordance with Rossi's (1971, 1978) results on the perception of Fo movements that are too short and/or too slight to be perceived as anything but level pitches. However, delimiting the beginning and end of a vowel (which determine the "2/3 point") is not a fool-proof procedure, and moving the point of measurement just 1 cs left or right may introduce quite considerable differences in the Fo measurement. Thus, the measures in table 4 should be taken cum grano salis.

For the first stressed vowel, [ˈɑ], two measures are given: its minimum value (which is turned into account in the calculation of the total rise of Fo in that, first stress group), and the value at 2/3 of the distance from the beginning of its rise (which is turned into account in fig. 7). From [ˈu] and [ˈi], and their time interval, a "base-line" slope, S_b, has been calculated, and, likewise, from [ˌʌ] and [ˌi], a "top-line" slope, S_t, is determined. The distance between each stressed and the immediately succeeding unstressed vowel (or sonorant consonant) is also given.

Note that stimuli have been grouped in table 4, as in fig. 3, according to the rank order of their responses, from maximally interrogative (rank 1), through maximally non-final, to maximally declarative (rank 15), as indicated in small script next to the stimulus-designations. The 6 groups of fig. 3 are maintained, and for each group averages have been calculated. Every measure has been assigned a rank, from highest (1) to lowest (15), as indicated in small raised script to the right of each value. Further, the averages have been assigned ranks in the same fashion, from highest (1) to lowest (6), as indicated in small lowered script.

If we let the "natural order" be the ranks assigned to stimuli according to responses, i.e. the order in which stimuli have been tabulated from top to bottom in table 4, then it is

Table 4 (continued on the next page)

Fundamental frequency measurements of segments and relations between certain segments in the 15 stimuli of test 1. Stimuli have been re-grouped according to identification by 10 Copenhagen speakers, cf. fig. 3. Numbers in small script indicate ranks from highest (^{1}) to lowest (^{15}) on a 15-point rank order scale. For an account of the measuring procedure, see the text.

	ˈɑ$_{min}$ (1)	ˈɑ$_{\frac{2}{3}}$ (2)	ŋ$_{max}$ (3)	ˈu	°ʌ	°ɑ	ˈi
I-1 ^{1}	$107^{10\frac{1}{2}}$	$116^{6\frac{1}{2}}$	162^{1}	$136^{1\frac{1}{2}}$	158^{1}	147^{1}	138^{1}
I-3 ^{2}	$112^{4\frac{1}{2}}$	$119^{1\frac{1}{2}}$	160^{2}	$136^{1\frac{1}{2}}$	156^{2}	141^{2}	134^{3}
I-2 ^{3}	111^{7}	112^{11}	153^{4}	135^{3}	150^{4}	139^{4}	137^{2}
Av. $_{1}$	110.0_{3}	115.7_{3}	158.3_{1}	135.7_{1}	154.7_{1}	142.3_{1}	136.3_{1}
I-5 ^{4}	104^{15}	$108^{13\frac{1}{2}}$	$148^{8\frac{1}{2}}$	125^{10}	$140^{10\frac{1}{2}}$	$134^{7\frac{1}{2}}$	126^{5}
I-4 ^{5}	$105^{13\frac{1}{2}}$	$108^{13\frac{1}{2}}$	140^{13}	$128^{6\frac{1}{2}}$	$142^{8\frac{1}{2}}$	$134^{7\frac{1}{2}}$	127^{4}
Av. $_{2}$	104.5_{6}	108.0_{6}	$144.0_{4\frac{1}{2}}$	126.5_{3}	141.0_{3}	134.0_{4}	126.5_{2}
NF-3 ^{6}	114^{2}	$119^{1\frac{1}{2}}$	$152^{5\frac{1}{2}}$	132^{4}	150^{4}	140^{3}	$124^{7\frac{1}{2}}$
NF-1 ^{7}	$112^{4\frac{1}{2}}$	$116^{6\frac{1}{2}}$	$152^{5\frac{1}{2}}$	130^{5}	150^{4}	$133^{10\frac{1}{2}}$	125^{6}
NF-2 ^{8}	$112^{4\frac{1}{2}}$	$118^{3\frac{1}{2}}$	150^{7}	$128^{6\frac{1}{2}}$	146^{6}	130^{12}	$124^{7\frac{1}{2}}$
Av. $_{3}$	112.7_{1}	117.7_{1}	151.7_{2}	130.0_{2}	148.7_{2}	$134.3_{2\frac{1}{2}}$	124.3_{3}
NF-5 $^{9}_{4}$	108^{9}_{4}	$110^{12}_{4\frac{1}{2}}$	$144^{11}_{4\frac{1}{2}}$	125^{10}_{4}	133^{14}_{5}	128^{14}_{5}	115^{9}_{4}
D-2 ^{10}	116^{1}	$118^{3\frac{1}{2}}$	144^{11}	126^{8}	143^{7}	136^{5}	113^{10}
D-3 ^{11}	110^{8}	113^{10}	$148^{8\frac{1}{2}}$	125^{10}	$140^{10\frac{1}{2}}$	$133^{10\frac{1}{2}}$	112^{11}
NF-4 ^{12}	$107^{10\frac{1}{2}}$	$115^{8\frac{1}{2}}$	144^{11}	$122^{12\frac{1}{2}}$	137^{12}	$134^{7\frac{1}{2}}$	$108^{12\frac{1}{2}}$
D-1 ^{13}	$112^{4\frac{1}{2}}$	117^{5}	154^{3}	119^{14}	$142^{8\frac{1}{2}}$	$134^{7\frac{1}{2}}$	105^{14}
Av. $_{5}$	111.3_{2}	115.8_{2}	147.5_{3}	123.0_{5}	140.5_{4}	$134.3_{2\frac{1}{2}}$	109.5_{5}
D-4 ^{14}	106^{12}	$115^{8\frac{1}{2}}$	134^{14}	$122^{12\frac{1}{2}}$	134^{13}	129^{13}	$108^{12\frac{1}{2}}$
D-5 ^{15}	$105^{13\frac{1}{2}}$	105^{15}	131^{15}	112^{15}	121^{15}	119^{15}	96^{15}
Av. $_{6}$	105.5_{5}	$110.0_{4\frac{1}{2}}$	132.5_{6}	117.0_{6}	127.5_{6}	124.0_{6}	102.0_{6}

1) ˈɑ$_{min}$ is measured at the turning point in the fall-rise of [ˈɑ].

2) ˈɑ$_{\frac{2}{3}}$ is measured at a point in time $\frac{2}{3}$ from the beginning of the rise of [ˈɑ].

3) ŋ$_{max}$ is the maximum value of [ŋ].

(continued from the preceding page) <u>Table 4</u>

˳i		S_b (4)		S_t (5)		ɑmin - ŋmax		ɑ$_3^2$ - ŋmax		ˈu - ˳ʌ		ˈi - ˳i			
153	1	-4.2	2	9.3	1	55	1	46	1	22	2	15	1	I-1	1
144	2½	4.3	5	23.5	6	48	2	41	2½	20	3½	15	6	I-3	2
144	2½	-4.4	1	11.5	3½	42	4½	41	2½	15	9½	7	11½	I-2	3
147.0	1	-1.4	1	14.8	3	48.3	1	42.7	1	19.0	1	10.7	2	Av.	1
134	6	-2.1	3	11.3	2	44	3	40	4	15	9½	8	9	I-5	4
134	6	2.0	4	15.4	5	35	12	32	10½	14	12	7	11½	I-4	5
134.0	2	0.0	2	13.4	2	39.5	2	36.0	2	14.5	4	7.5	5	Av.	2
137	4	17.0	8	24.0	7	38	8	33	9	18	5½	13	2	NF-3	6
134	6	10.4	7	29.1	8	40	6	36	6	20	3½	9	7	NF-1	7
130	8	8.5	6	30.8	9	38	8	32	10½	18	5½	6	13	NF-2	8
133.7	3	12.0	3	28.0	4	38.7	3	33.7	4	18.7	2	9.3	3	Av.	3
127	9 / 4	20.8	9 / 4	11.5	3½ / 1	36	11 / 5	34	8 / 3	8	15 / 6	12	3½ / 1	NF-5	9 / 4
124	10	26.5	10	36.5	11	28	13½	26	13½	17	7	11	5	D-2	10
115	12	27.7	11	45.5	13	38	8	35	7	15	9½	3	15	D-3	11
120	11	29.8	13	33.3	10	37	10	29	12	15	9½	12	3½	NF-4	12
113	13½	30.4	14	54.7	14	42	4½	37	5	23	1	8	9	D-1	13
118.0	5	28.6	5	42.5	5	36.3	4	31.8	5	17.5	3	8.5	4	Av.	5
113	13½	29.2	12	45.3	12	28	13½	19	15	12	13	5	14	D-4	14
104	15	34.0	15	70.4	15	26	15	26	13½	9	14	5	9	D-5	15
108.5	6	31.6	6	57.9	6	27.0	6	22.5	6	10.5	5	6.5	6	Av.	6

4) S_b is the slope of the intonation contour ("base-line" slope), in cs/sec, based on the measurements of [ˈu] and [ˈi], and their time interval.

5) S_t is the slope of the "top-line", in cs/sec, based on the measurements of [˳ʌ] and [˳i], and their time interval.

immediately obvious that certain parameters deviate less from this natural order than do others. And if we just look at the ranks of the averages, S_b, [ˈi] and [ˌi] are even seen to follow the natural order exactly, and the rank order for the averages of [ˈu] is reversed in only one place. These parameters, therefore, look like good candidates and worth inspecting more closely, in the search for correlations between the physical properties of stimuli and the responses they yielded. - Fig. 7 is a stylized graph of the intonation contours and the final rise from stressed to unstressed [i], based on the measures of [ˈu], [ˈi] and [ˌi]. To complete the picture of the contour, the value of ˈɑ$^2_3$ has been included (plus 15% to compensate for intrinsic Fo level differences between [ˈu] and [ˈi], on the one hand, and [ˈɑ] on the other, for this particular speaker, cf. Reinholt Petersen, 1976).

Fig. 7 clarifies figs. 5 and 6 and "explains" some things that were less transparent in those figures: for instance, I-4 and I-5 were seen in fig. 6 to be very nearly concurrent with NF-1, and one was left to wonder why, then, had these two interrogative sentences not been labelled "non-final" more than about 50% of the time by listeners? - It is clear from fig. 7 that the slopes of I-4,5 are different from those of the well identified non-finals, i.e. they are very nearly flat, whereas the slopes of NF-1,2,3 are falling. - In short, the six groups of fig. 3 and table 4 emerge rather clearly.

Since the members from each group do not differ significantly among themselves, as far as distribution of responses goes, one might leave it at that and be satisfied that slope, final stressed and final unstressed vowel all correlate exactly with responses when averages are considered. - However, this is slightly unsatisfactory, partly because stimuli <u>are</u> different, and <u>did</u> receive different responses, partly because it is not possible to say which of the three parameters is the most crucial for identification. - Accordingly, the data in table 4 were turned into account for the calculation of correlations between responses, on one hand, and each of the parameters of table 4, on the other.

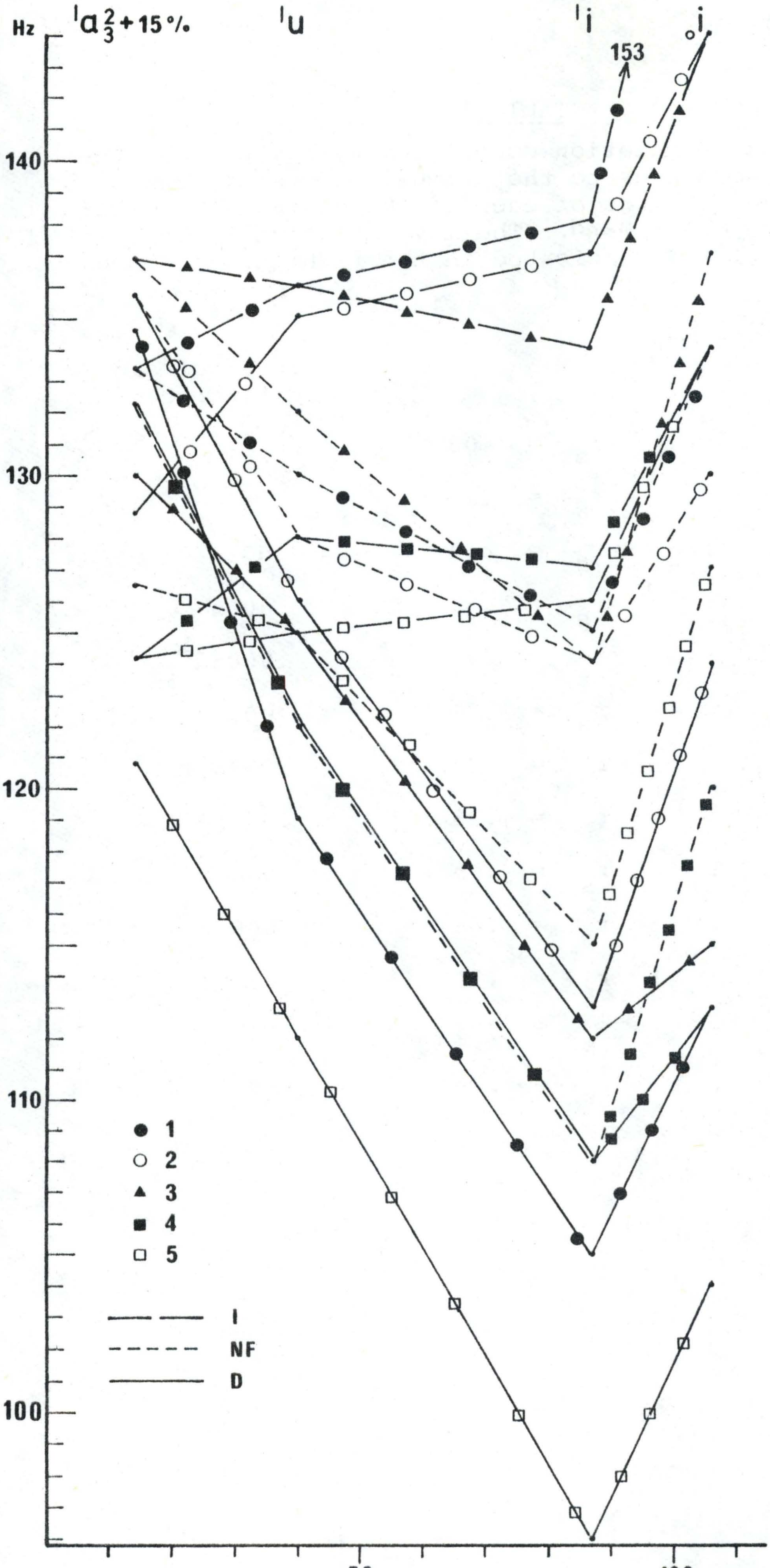


Figure 7

Stylized intonation contours plus the final rise from stressed to unstressed [i] of the 15 stimuli in test 1. See further the text.

Table 5

Spearman rank correlation coefficients, r_s, between the rankings of responses to the stimuli of test 1, on one hand, and the rankings of each of the parameters in table 4, on the other hand. The level of significance (if better than .10) is indicated in the right hand column.

Correlation between response rankings and rankings of			level of significance
ˈɑ$_{min}$	:	r_s = .08	-
ˈɑ$_{\frac{2}{3}}$	:	r_s = .18	-
ɔ$_{max}$	:	r_s = .64	.005
ˈu	:	r_s = .89	.0005
˳ʌ	:	r_s = .77	.0005
˳ɑ	:	r_s = .71	.005
ˈi	:	r_s = .98	.0005
˳i	:	r_s = .98	.0005
S_b	:	r_s = .95	.0005
S_t	:	r_s = .88	.0005
ɑ$_{min}$-ɔ$_{max}$	:	r_s = .71	.005
ɑ$_{\frac{2}{3}}$-ɔ$_{max}$	:	r_s = .74	.005
ˈu-˳ʌ	:	r_s = .38	.10
ˈi-˳i	:	r_s = .30	-

Table 5 gives Spearman rank correlation coefficients (calculated, when necessary, with corrections for tied ranks) between the rankings of responses to the stimuli, and the rankings of the measures of various points in the fundamental frequency trancings, as presented in table 4.

It is obvious from table 5 that S_b, ˈi, and ˌi, i.e. the slope of the intonation contour (or "base-line" slope), the levels of the final stressed and the final unstressed vowel, respectively, all correlate very highly with responses.

2.4.1.2 Terminal rises

Before we proceed, it is interesting to note that the magnitude of the final rise, i.e. ˈi−ˌi, shows only a slight and non-significant correlation with responses (and no correlation at all (r_s = .05) when this rise is expressed in percentage of the level of the stressed vowel). This must be taken to mean that the magnitude of the final rise was not, in this experiment, a perceptual cue to the identification of intonation contours, but its placement on the frequency scale is (i.e. the levels of the final stressed and succeeding unstressed vowels).

However, Danish listeners are not insensitive to terminal rises: In the course of some informal experiments, conducted at the Institute of Linguistics, Uppsala University, with the ILS-system for analysis and synthesis of speech, the declarative sentence "Der går mange busser fra Tiflis." was recorded (by the author). It was then re-synthesized with the final unstressed syllable, "-lis", being raised 60Hz, in steps of 10Hz (everything else being kept constant). These 6 utterances, together with the original, were later played back to a group of phoneticians in Copenhagen. The first two steps up of "-lis" made no apparent difference. At +30Hz and +40Hz listeners reported that they heard a statement, but a rather obstinate one. At +50Hz the utterance began to sound interrogative (to some listeners), and at 60Hz the utterance was classified as a question by the majority, but a rather surprised or disbelieving echo-question. Other listeners would not accept the utterance as a question at all, but dismissed it as being wholly unnatural.

This illustrates very nicely a point where real and synthetic speech part company: it is possible to make listeners respond in certain (and often predictable) ways to cues, or rather combinations of cues (cf. below), which are not found in real speech, but analyses of and perceptual experiments on real speech may then serve to dissolve any ambiguity as to which of the cues, found to be operative in synthetic speech, is the dominant one in speech perception. The interpretation presented here for Danish could probably also be applied to the results obtained for Swedish (and American): Hadding-Koch (1961) found in her analysis of Swedish that statements tend to end in a terminal fall, questions in a terminal rise (but there were differences to observe in earlier parts of the utterance, also). In experiments with synthetic stimuli Studdert-Kennedy and Hadding (1972, 1973) found that if a low peak (in the stressed vowel of 'November') is heard, listeners tend to interpret the utterance as a statement unless it is followed by a large terminal rise (in the final, unstressed syllable). If a high peak is heard, listeners tend to interpret the utterance as a question, unless they also hear a large terminal fall. In other words: everything else being equal, the higher the terminal rise, the more questions are heard, and vice versa, or, again, everything else being equal, the higher the terminal peak, the more questions are heard, and vice versa. They conclude that peak and perceived terminal glide are the two factors that seem to govern linguistic judgments of intonation contours (although they are not just and simply additive in their effects). However, in consequence of the line of argument for Danish above, I would be inclined to give top priority to their peak level.

2.4.2 Correlations and concordances between various points in the course of fundamental frequency in the stimuli

Although table 5 is informative, it may also be slightly misleading. The high correlations we observe between responses and S_b, [ˈi], and [ˌi] (and [ˈu] and S_t), respectively, are of course partly due to the fact that these parameters are intercorrelated themselves. If [ˈu] is high, [ˈi] is likely to be high also, etc.; in fact, this is precisely the kind of prediction that can be read off from the model in fig. 1.

The Spearman rank correlation coefficients (with corrections for tied observations, when necessary) between a number of pairs of measurements in the tracings, as well as the Kendall coefficient of concordance between several measures, have been calculated and are presented in table 6. The level of significance, when better than .10, is indicated in the right hand column.

Table 6 focuses upon the maxima (first unstressed syllable in each stress group) and minima (the stressed syllables) and their correlations to each other. A couple of aspects seem worth commenting upon. - If we set the limit between high and low correlations at $r_s = .70$, we note, first of all, that all correlations involving the first stressed vowel, $ˈɑ^2_3$ or $ˈɑ_{min}$[1], are low and so are, with one exception ($ŋ_{max} \sim {}_\circ ʌ$) all correlations involving $ŋ_{max}$. The course of fundamental frequency during the first stress group, in short, does not correlate very highly with later parts of the utterance.

Correlations between the stressed vowel and the rise to the succeeding unstressed vowel, in the second and third stress groups, are also low. Apart from that, every point of measurement in the second and third stress groups correlate highly with every other point, and "top-line" and "base-line" slopes, S_t and S_b, are highly correlated. (Further, the stressed vowels correlate highly with S_b and the unstressed vowels with S_t - but one should bear in mind that these slopes have been calculated from [ˈu] and [ˈi], and [˳ʌ] and [˳i], respectively, and are thus anything but independent of these measures.)

The degree of concordance between the three maxima (unstressed syllables) is considerable and is higher than that between the

---

1) $ˈɑ_{min}$ renders lower coefficients than does $ˈɑ^2_3$, and since $ˈɑ^2_3$ thus seems to be the more "consistent" of the two measurements, correlation coefficients between $ˈɑ_{min}$ and <u>all</u> other points have not been calculated.

<u>Table 6</u> (continued on the next page)

Spearman rank correlation coefficients, r_s, between the rankings of a number of pairs of parameters in table 4, and Kendall coefficients of concordance, W, between the rankings of several such parameters. The level of significance (if better than .10) is indicated in the right hand column.

				level of significance
'i ~ S_b	:	r_s =	.98	.0005
'i ~ 'u	:	r_s =	.92	.0005
'i ~ α_{min}	:	r_s =	.01	-
'i ~ 'α_3^2	:	r_s =	.11	-
'i ~ $_\circ$i	:	r_s =	.96	.0005
'i ~ $_\circ$∧	:	r_s =	.75	.005
'i ~ ŋ$_{max}$	:	r_s =	.57	.025
'i ~ 'i - $_\circ$i	:	r_s =	.19	-
'u ~ S_b	:	r_s =	.82	.0005
'u ~ 'α_{min}	:	r_s =	.38	.10
'u ~ 'α_3^2	:	r_s =	.45	.05
'u ~ $_\circ$∧	:	r_s =	.90	.0005
'u ~ $_\circ$i	:	r_s =	.94	.0005
'u ~ ŋ$_{max}$	:	r_s =	.67	.005
'u ~ 'u - $_\circ$∧	:	r_s =	.48	.05

Table 6 (continued from preceding page)

				level of significance
$'\alpha^2_3 \sim S_b$	:	r_s	= .02	-
$'\alpha^2_3 \sim \eta_{max}$	:	r_s	= .58	.025
$'\alpha^2_3 \sim {}_\circ\Lambda$	:	r_s	= .66	.005
$'\alpha^2_3 \sim {}_\circ i$	:	r_s	= .28	-
$'\alpha^2_3 \sim {}'\alpha^2_3 - \eta_{max}$	:	r_s	= .12	-
$'\alpha_{min} \sim \eta_{max}$	:	r_s	= .49	.05
$'\alpha_{min} \sim {}'\alpha_{min} - \eta_{max}$	:	r_s	= .34	-
$\eta_{max} \sim {}_\circ\Lambda$	:	r_s	= .84	.0005
$\eta_{max} \sim {}_\circ i$	:	r_s	= .66	.005
${}_\circ\Lambda \sim {}_\circ i$	:	r_s	= .82	.0005
$S_b \sim S_t$	:	r_s	= .89	.0005
$\eta_{max} \sim {}_\circ\Lambda \sim {}_\circ i$	:	W	= .84	.005
$'\alpha^2_3 \sim {}'u \sim {}'i$	:	W	= .66	.01
$'\alpha_{min} \sim {}'u \sim {}'i$	:	W	= .61	.025
$'\alpha^2_3 \sim \eta_{max} \sim {}'u \sim {}_\circ\Lambda \sim {}'i \sim {}_\circ i$	:	W	= .73	.0005
$'\alpha_{min} \sim \eta_{max} \sim {}'u \sim {}_\circ\Lambda \sim {}'i \sim {}_\circ i$	:	W	= .70	.0005

three minima (stressed vowels), which is somewhat surprising, - but it might possibly reflect a greater uncertainty about measuring Fo in [ˈɑ], i.e. finding the relevant point to measure, than in other vowels. Finally, the concordance between all six points, maxima and minima, is high as well. (Again, when $ˈɑ^2_3$ is used, the concordance is slightly higher than when $ˈɑ_{min}$ is used.)

The fact that the two stressed vowels, [ˈu] and [ˈi], correlate highly with [ˌʌ] and [ˌi], respectively, but not with the magnitude of the rise from stressed to unstressed vowel, together with the fact that responses show low correlations with these rises, cf. table 5, may be interpreted thus: the distance between stressed and first unstressed vowel in a stress group, which decreases from beginning to end of the utterance, does not, also, decrease as a function of the kind of intonation contour upon which the stress group patterns are imposed. In other words, top-lines and base-lines decline (more or less, depending on the sentence) and top-lines decline more rapidly than do base-lines. But top-lines do not decline relatively more rapidly, the steeper the base-line is:

The differences between S_t and S_b were calculated and ranked. The correlation coefficient between this ranking and that of S_b is .18 only, - whereas it is .89 between the rankings of S_b and S_t. - Thus, base-line and top-line are highly correlated, and top-lines are steeper than base-lines, cf. table 4, but the difference in slope between top- and base-lines is, within certain limits, more or less random, averaging about 15Hz/sec. - However, only a subset of the recordings of one subject is involved, and nothing definite can be said about this point yet.

Since all points in the course of fundamental frequency in the second and third stress groups correlate rather highly with each other, the high correlations listed in table 5 between responses and [ˈi], [ˌi], S_b, [ˈu], and S_t cannot be said to be mutually independent. On the contrary, there is good reason to believe that identification has been based on the whole Fo course during the second and third stress groups, rather than on one single parameter, but, as mentioned in the abstract, the level of the final stressed and unstressed vowels can almost, alone, account for the way responses were distributed over the 15 stimuli.

One example will suffice to illustrate the point made above: [ˈi] and [ˌi] gave the highest correlation with responses, and they correlate highly with each other as well. The Kendall rank correlation coefficient, τ, between response and [ˈi] is .93, between response and [ˌi] it is .94, and between [ˈi] and [ˌi] it is .91.

The Kendall partial rank correlation coefficient between response and [ˈi], with the effect of $[_{\circ}i]$ partialled out, is .57, and the partial rank correlation between response and $[_{\circ}i]$, with the effect of [ˈi] partialled out, is .60. Both coefficients are a good deal lower than the non-partial correlations, and we may conclude that the relation between response and [ˈi] is *not* independent of $[_{\circ}i]$, and, vice versa, the relation between response and $[_{\circ}i]$ is *not* independent of [ˈi]. Since the non-partial correlation coefficients decrease by approximately the same amount we cannot determine which of the two, [ˈi] or $[_{\circ}i]$, is more dependent upon the other.

2.5 Conclusion - test 1

In conclusion we may say about the results of test 1, that they confirm the implications of the model in fig. 1 and the working hypothesis outlined in section 2.4.1: intonation contours are more widely separated in later than in early parts of the utterance, and the stressed syllables in this later part are more decisive for identification than are the unstressed syllables, cf. table 5. Since the correlation/concordance between points in the Fo course in the second and third stress groups is high, cf. table 6, it is not possible to establish a single parameter which, independently of all other parameters, will account for listeners' identification. Rather, identification may be determined by all of the later part of the Fo course, probably, however, with the *level* of the final stress group having slightly more weight than that of the second stress group. The terminal *rise* from stressed to unstressed syllable was of no consequence whatsoever.

Copenhagen speakers identified three clear categories, non-Copenhagen speakers did not. - Whether these three categories are purely phonetic, i.e. a matter of dividing a physical (near-) continuum into X number of classes, or whether this categorization reflects some real linguistic categories is a matter for test 2 to resolve. - However, I did think that there was some support for a hypothesis of linguistic categories already in the fact that speakers do indeed produce more than two kinds of intonation contours, which do tend to arrange themselves along three slopes, with a good deal of overlap, however, between individual items.

3. Test 2

If the categorization performed by subjects in test 1 were purely phonetic, or non-linguistic, i.e. the continuum had been divided into three equally large bits, without regard to any linguistic analysis, then one would expect listeners to divide the same continuum into two equally large bits, when presented with only two categories to choose from. Accordingly, the test tape from test 1 was presented to 7 of the 10 Copenhagen speakers (and two of the four non-Copenhagen speakers), about 6 months later. The procedure resembled exactly that of test 1, except that stimuli now were to be labelled either "declarative" or "non-declarative". (Subjects did not know that the two tapes were identical.)

3.1 Results - test 2

Table 7 and fig. 8 present the 7 Copenhagen speakers' distribution of responses in test 1 to the 15 stimuli. Comparing fig. 8 to fig. 3, we see that the response distributions are almost identical: the same 6 groups turn up in fig. 8 as in fig. 3, but the ordering of stimuli within some groups is slightly different. Thus, I-3 and I-1, I-4 and I-5, NF-1 and NF-3, and NF-4 and D-2 have changed places. However, the Spearman rank correlation coefficient between response rankings by 10 and by 7 listeners, respectively, is .99, and we may safely conclude that the 7 listeners employed in test 2 are a representative subset of the 10 listeners employed in test 1.

Table 8 and fig. 9 present the results of test 2. As in test 1, there was no "order-effect" to be found with any subject, nor did subjects differ significantly among themselves.

It is evident that stimuli have not, in test 2, been divided into two equally large classes. There are differences in the rank ordering of stimuli in test 2 and test 1 (for the 7 listeners) but the Spearman rank correlation coefficient between these rankings is .96 and the response distributions in the two tests are thus highly correlated. The only non-predictable difference between figs. 8 and 9 is that whereas the cross-over point between the non-final and declarative categories lies to the right of stimulus NF-5 in fig. 8, the corresponding cross-over point between the non-declarative and declarative categories lies to the left of NF-5 in fig. 9. But the cross-over point is still sharp and well-defined, and only three groups are now formed, cf. the parentheses around stimuli designations in fig. 9. (A group is formed by stimuli whose response distributions cannot, by a χ^2 test, be proved to be significantly different at the .10 level.)

About the change of NF-5 from one category to another in the two tests: in fact, its identification in test 2 as mainly declarative (67% of the time) is less curious than its identification in test 2 as mainly non-final (79% of the time by the 7 Copenhagen speakers). If we go back to (table 4 and) fig. 7, we see that NF-5, in slope and position on the frequency scale, runs very close to those stimuli that have been identified mainly as declarative (D-1,2,3,4,5 and NF-4). This would have warranted an identification in test 1 as declarative. However, it <u>is</u> the "top" contour in that group of "low" ones in fig. 7, and its identification as mainly non-final in test 1 might possibly reflect an attempt by listeners to balance responses (i.e. to try and give 33% responses in each category) - and if one of the "low" contours is to be affected by this attempt, it is precisely NF-5.

The two non-Copenhagen speakers who took test 2 did not, this time, differ remarkably from the Copenhagen speakers. The cross-over point between the two categories is in the same place as for Copenhagen speakers. But to the left of NF-5 everything is unanimously labelled "non-declarative", whereas the identification functions to the right of the cross-over point are slightly more slowly rising and falling than for Copenhagen speakers. Again, however, 4 and 2 (non-homogeneous) non-Copenhagen speakers, respectively, are not sufficient for sound conclusions to be drawn, nor for attempts at explaining their behaviour.

In conclusion, the stimuli that were labelled "non-final" in test 1 were <u>not</u> split up into two equally sized bits, each going to the declarative and non-declarative categories, respectively. With one exception, the formerly "non-final" stimuli were now categorized as "non-declarative" and this may be taken as an indication that linguistic categories were at play. (A further support

Table 7

7 Copenhagen speakers' response to the 15 stimuli in test 1.

response / stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	61	9		70
D-2	52	18		70
D-3	56	14		70
D-4	70			70
D-5	70			70
NF-1		59	11	70
NF-2	3	63	4	70
NF-3	1	58	11	70
NF-4	56	14		70
NF-5	20	50		70
I-1		3	67	70
I-2		4	66	70
I-3		1	69	70
I-4		32	38	70
I-5		34	36	70
Total	389	359	302	1050

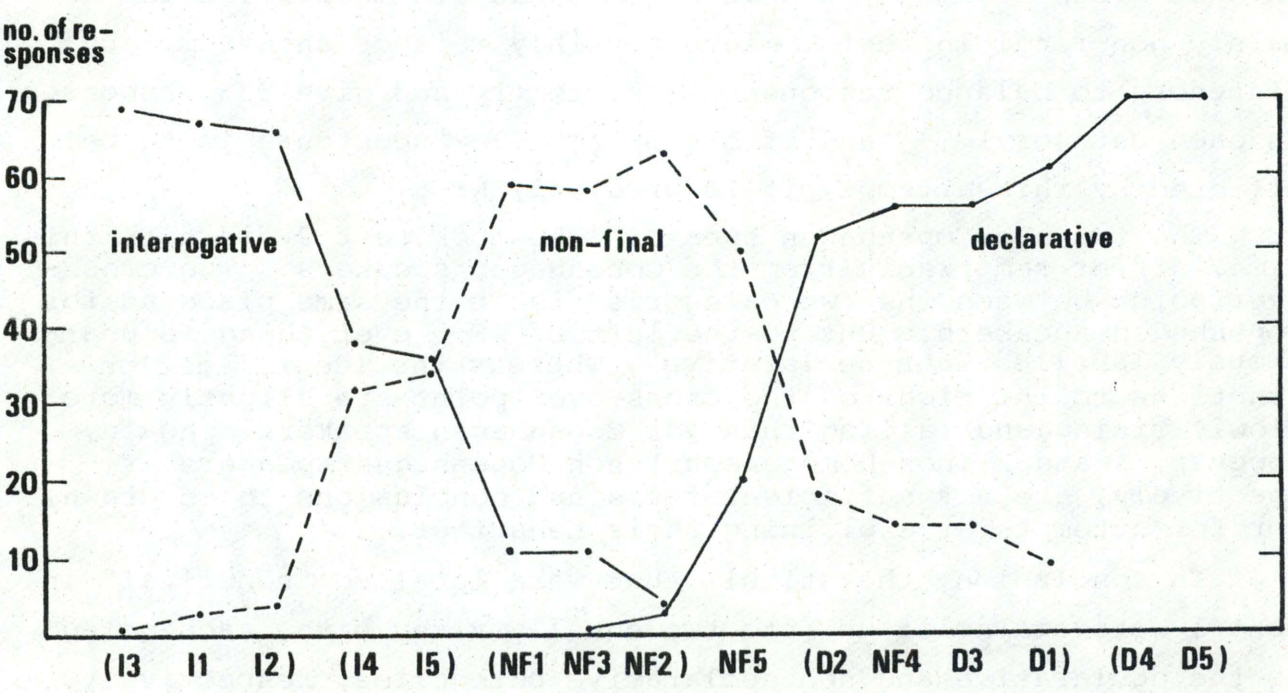


Figure 8

Identification functions for the 15 stimuli in test 1, as identified by 7 Copenhagen speakers.

Table 8

7 Copenhagen speakers' response to the 15 stimuli in test 2.

response / stimulus	DECLARATIVE	NON-DECLARATIVE	Total
D-1	68	2	70
D-2	66	4	70
D-3	63	7	70
D-4	70		70
D-5	70		70
NF-1		70	70
NF-2	8	62	70
NF-3	3	67	70
NF-4	59	11	70
NF-5	47	23	70
I-1		70	70
I-2		70	70
I-3		70	70
I-4		70	70
I-5	1	69	70
Total	455	595	1050

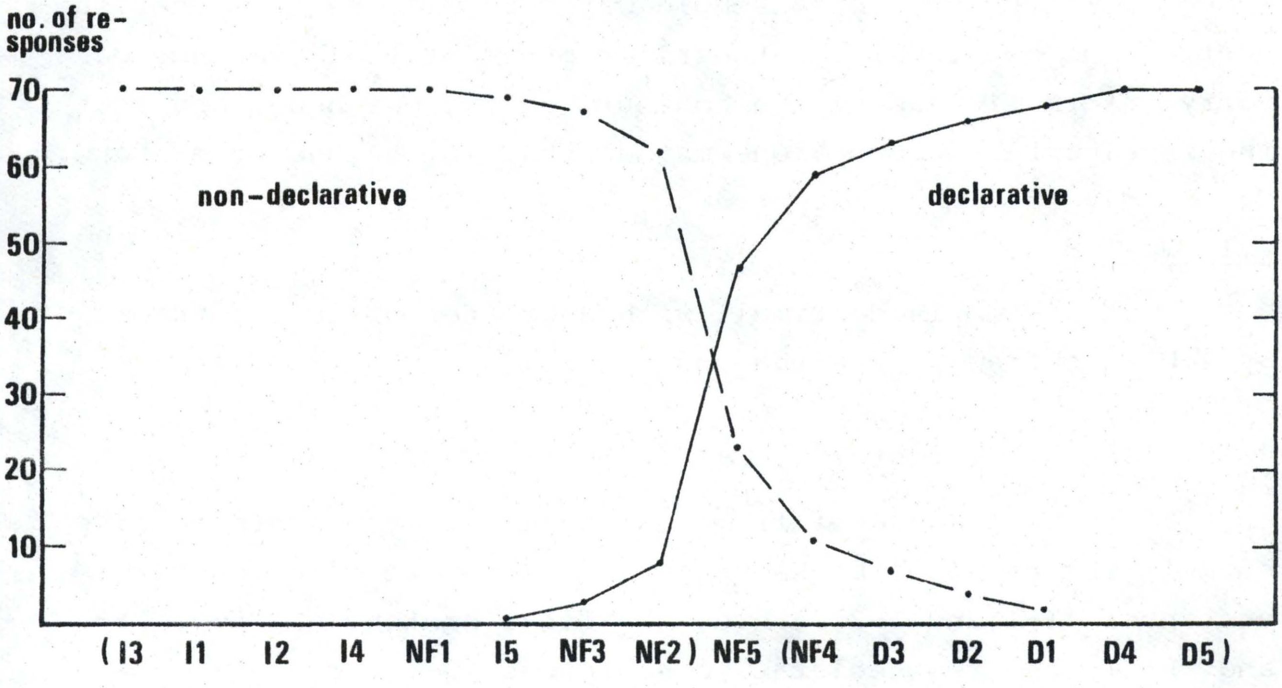


Figure 9

Identification functions for the 15 stimuli in test 2, as identified by 7 Copenhagen speakers.

of this hypothesis would have been if running the same test, but with response labels "interrogative" and "non-interrogative", had given a cross-over point in the identification functions in the neighbourhood of stimuli I-4/NF-3, cf. fig. 3. This ought, as a matter of fact, to be tried.)

4. Tests 3 and 4

Tests 3 and 4 were designed to answer the third question in the introduction: how early or how late in the utterances are the various contours perceptually differentiated? From the results of test 1 (table 5, more precisely), we might expect stimuli to be rather well identified even when the third stress group is missing, given that responses to test 1 correlated fairly highly with [ˈu] and [ˌʌ], - but identification should not be much better than chance, given only the first stress group to judge from. Likewise, we may predict that identification should not be affected much in the absence of the first and second stress groups. If the predictions hold water it will be a further support of the hypothesis that earlier parts of the utterance do carry information about the contour, on the one hand, and that the final stress group alone may do the job, on the other hand.

4.1 Test 3

Test 3 comprised stimuli with a greater or lesser number of syllables cut away from the end.

4.1.1 Stimuli - test 3

7 of the 15 stimuli from test 1 served as parents for the stimuli in test 3. The seven stimuli were two interrogative sentences (I-1 and I-2), three non-final sentences (NF-1, NF-2, and NF-3), and two declarative sentences (D-1 and D-4) - all stimuli that had been well identified by the Copenhagen speakers in test 1, cf. figs. 3 and 8. Three non-finals were included

because this was the category which, although well identified, subjects had been least sure of. D-5 was omitted because it was so very much different from the other declarative sentences, cf. figs. 1 and 7, and in a test with relatively few stimuli no one stimulus ought to stand out clearly from the rest. - The rationale behind choosing, for test 3, stimuli that had been well identified in test 1 was simple: stimuli that had been ambiguous, given the whole of three stress groups to judge from, would be even more so when they were mutilated. However, as Robert Porter has pointed out to me in a personal communication, this may be an erroneous assumption. When stimuli are mutilated, identification criteria must necessarily change, - and originally ambiguous stimuli might turn out to reveal some interesting features. (However, even with only 7 parent stimuli, the test was already rather long - but these 7 might, of course, still have included one or two of the originally ambiguous stimuli.)

By means of a segmentator, the 7 parent stimuli were being relieved step-wise of the last 5 syllables, thus:

cut no. 1: mange busser fra Tif

cut no. 2: mange busser fra

cut no. 3: mange busser

cut no. 4: mange bus

cut no. 5: mange

7 parents times 5 cuts give a total of 35 stimuli. Two test tapes, each containing 2x35 randomized stimuli, plus 5 introductory dummies, i.e. two tapes with 80, differently randomized, stimuli on each were prepared. Stimuli were introduced by numbers (read by the author) with a 3 sec interval for responding. The total duration of one run of each test tape was 8 minutes.

Subjects were instructed exactly as in test 1, cf. section 2.2.

One test tape was presented three times, the other twice, giving a total of 10 responses to each stimulus by each subject.

No subject did more than one run of a test tape per day. All (four) subjects listened over head-phones.

4.1.2 Subjects - test 3

4 of the 10 Copenhagen speakers (all phoneticians) from test 1 served as subjects in test 3. Four is, certainly, a small number, and the results presented below thus may not be conclusive, but they may at least serve as indications of certain trends.

4.1.3 Results - test 3

As in tests 1 and 2, there is no "order-effect" to be detected, and subjects do not differ significantly among themselves.

Fig. 10, top left, and table 9 present the 4 subjects' distribution of responses in test 1 to the 7 parent stimuli employed in test 3. Compared to fig. 3, NF-1 and NF-3 have changed places (still being members of the same group), but the Spearman rank correlation coefficient between the rankings of responses to the 7 stimuli used in test 3, by 10 subjects (fig. 3), and the ranking of responses by four subjects to these stimuli (fig. 10, top left) is .96 - and as for the 7 subjects in test 2, we may conclude that the 4 subjects in test 3 are a representative subset of the 10 subjects employed in test 1.

Fig. 10 and table 10 present the results. (In fig. 10, stimuli have been grouped according to the same criteria as in figs. 3, 4, 8 and 9, cf. sections 2.4 and 3.1.)

4.1.3.1 Cut no. 1: "mange busser fra Tif" - test 3

Fig. 10, top right, shows three clear peaks in the identification functions. Comparison with response distributions to the parent stimuli, fig. 10 top left (and tables 9 and 10), shows the same order of stimuli, but a raised proportion of "declarative"-responses, at the expense of both "interrogative" and "non-final" responses. And a χ^2 test on the distributions of declarative-responses in the parent and cut no. 1 cases shows

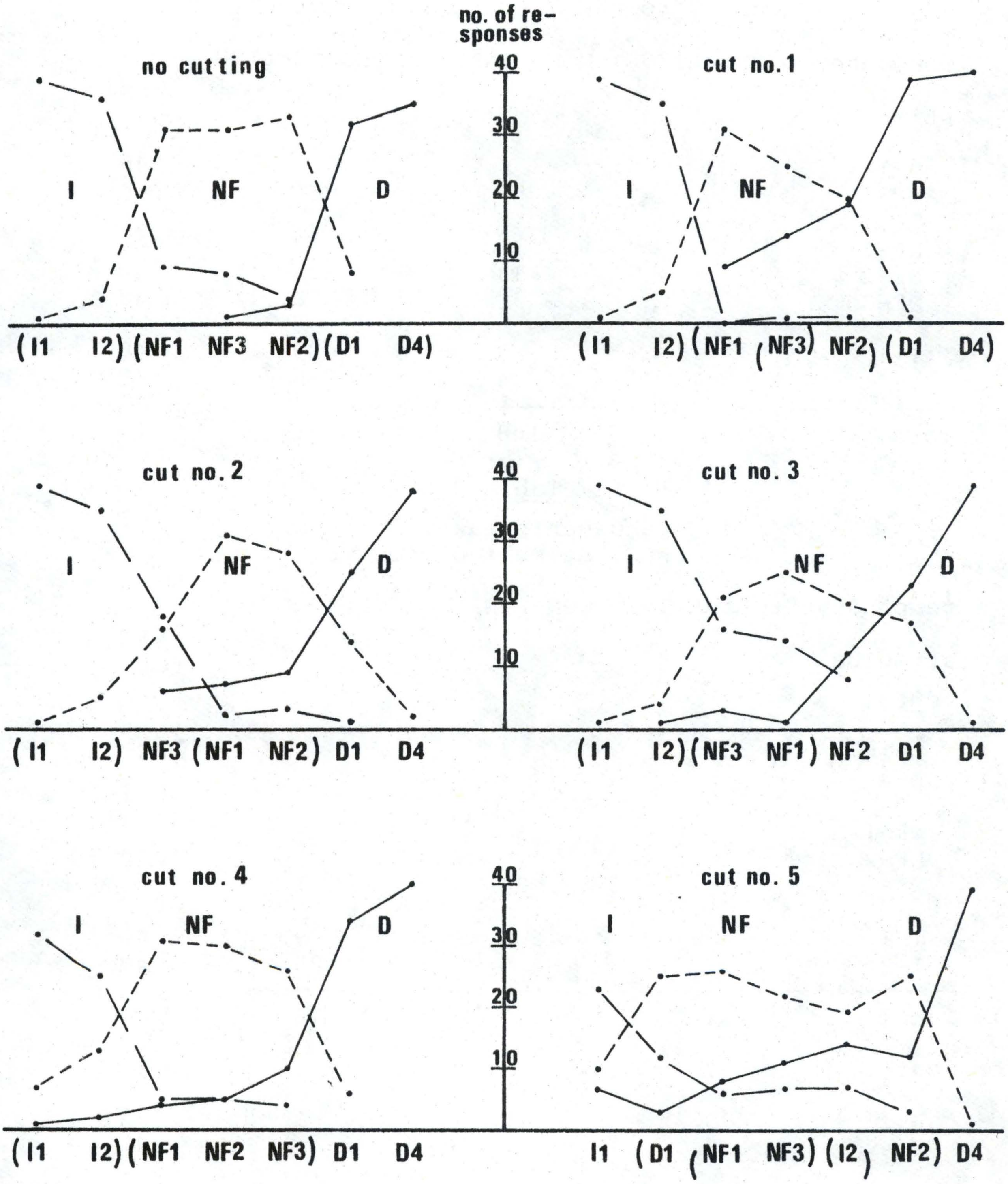


Figure 10

Identification functions for the 7 stimuli in test 3, as identified by 4 Copenhagen speakers.

Table 9

4 Copenhagen speakers' response in test 1 to the 7 stimuli of test 3.

response / stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	32	8		40
D-4	40			40
NF-1		31	9	40
NF-2	3	33	4	40
NF-3	1	31	8	40
I-1		1	39	40
I-2		4	36	40
Total	76	108	96	280

Table 10

4 Copenhagen speakers' response to the 7 stimuli of test 3.

response / stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
cut 1				
D-1	39	1		40
D-4	40			40
NF-1	9	31		40
NF-2	19	20	1	40
NF-3	14	25	1	40
I-1		1	39	40
I-2		5	35	40
Total	121	83	76	280
cut 2				
D-1	25	14	1	40
D-4	38	2		40
NF-1	7	31	2	40
NF-2	9	38	3	40
NF-3	6	16	18	40
I-1		1	39	40
I-2		5	35	40
Total	85	97	98	280

(continued on the next page)

Table 10 (continued from the preceding page)

4 Copenhagen speakers' response to the 7 stimuli of test 3.

response / stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
cut 3				
D-1	23	17		40
D-4	39	1		40
NF-1	1	25	14	40
NF-2	12	20	8	40
NF-3	3	21	16	40
I-1		1	39	40
I-2	1	4	35	40
Total	79	89	112	280
cut 4				
D-1	34	6		40
D-4	40			40
NF-1	4	31	5	40
NF-2	5	30	5	40
NF-3	10	26	4	40
I-1	1	7	32	40
I-2	2	13	25	40
Total	96	113	71	280
cut 5				
D-1	3	25	12	40
D-4	39	1		40
NF-1	8	26	6	40
NF-2	12	25	3	40
NF-3	11	22	7	40
I-1	7	10	23	40
I-2	14	19	7	40
Total	94	128	58	280

that this difference is statistically significant beyond the .001 level.

Losing the last unstressed syllable thus pulls in the direction of declarative-identification. The easy explanation would be that the terminal movement is, after all, essential and, everything else being equal, the lack of a final rise increases the number of declarative-identifications. - However, this is too facile: It is important to note that as soon as one cuts away bits and pieces, the utterance is no longer "natural" and does not necessarily resemble and contain the same amount of information as a corresponding non-mutilated one, in this case one that never had a final unstressed syllable (like: "mange busser fra Brest", for instance). In fact, another part of the material for the original analysis (Thorsen, 1976) contained questions and non-final periods, all variations on: ".. mange timer i statistik.." (the final syllable in "statistik" is stressed). And although the level of this final stressed syllable, "-stik", varies according to sentence type in the same fashion as does "Tif-", there are also minor differences to be observed in the movement of Fo within the stressed [i]: the higher it is on the frequency scale, the less falling it also is (and in some instances, with some subjects, it is even slightly rising, - but this is the exception rather than the rule). This variation in Fo movement is not, generally, found in the stressed vowel of "Tiflis". Thus, if variation in Fo movement is a general feature of stress groups consisting only of one (stressed) syllable, it may be interpreted as a compensation, in the absence of the high rise that a succeeding unstressed syllable would have brought about.

A more appealing account of the discrepancy (which is, after all, not very great) between fig. 10 top right and top left would therefore be that the Fo movement of the vowel in "Tif-" in NF-3 and NF-2 (which are those two stimuli that contribute most to the difference between response distributions) are too steeply falling (cf. fig. 2) to perceptually compensate for the lack of a final unstressed syllable, wherefore NF-3 and NF-2 begin to resemble declarative utterances. (But this presupposes that subjects judged the stimuli in cut no. 1 as if it were a complete utterance.) As it turns out, the pattern almost repeats itself when we get to cuts no. 3 and 4 (mange busser vs. mange bus), and I think the explanation just offered is a fairly likely one. (See, however, a general reservation towards this comparison between parents and offspring at the end of section 4.2.5.)

In table 11 are given Spearman rank correlation coefficients between rankings of responses by the four subjects to the stimuli in test 3 and rankings of a number of the Fo measurements presented in table 4. Since cuts no. 2 and no. 3 had the same response rankings, they are treated together. Note that the correlation coefficients in table 11 are not directly comparable to those in table 5 which are based on rankings of 15 stimuli, whereas table 11 is based on rankings of 7 stimuli. But the coefficients in table 11 are comparable within and across cuts.

In cut no. 1 [ˈi] and [ˌʌ] are seen to correlate highly with responses, with [ˈu] as a close follow-up. [ˌi] is no longer there, and [ˌʌ] seems to have taken over its role, being now the last maximum in the utterance.

4.1.3.2 Cut no. 2: "mange busser fra" - test 3

It is apparent from fig. 10, mid left, that losing the whole of the third stress group is not detrimental to the identification of three categories, - and, further, we are back at a distribution which resembles that of the identification of parent stimuli rather closely, but the order of stimuli has been shifted in one place: NF-3 is moved one step to the left in cut no. 2. However, the identification functions are slightly less steep at the crossover points than in fig. 10, top left.

Table 11 says that [ˈu] and [ˌʌ] are most highly correlated with responses (and [ˈɑ] is closely following them), and they have evidently taken over the role played by [ˈi] and [ˌi] in the parent stimuli, being now the last stress group in the stimuli. (Note that this cut ends in a terminal fall (as does any stress group that contains more than one post-tonic syllable) cf. fig. 2, but this does not diminish the number of interrogative-responses.)

Table 11 (continued on the following page)

Spearman rank correlation coefficients, r_s, between the rankings of responses to the stimuli of test 3, on one hand, and the rankings of certain Fo measurements in these stimuli, cf. table 4. The level of significance (if better than .10) is indicated in the right hand column.

Correlation between response rankings and rankings of			level of significance
Cut no. 1	ˈɑ$_{min}$	: r_s = 0	
	ˈɑ$^{2}_{3}$	: r_s = -.21	-
	ŋ$_{max}$	: r_s = .63	.10
	ˈu	: r_s = .93	.005
	$_{o}$ʌ	: r_s = .96	.0005
	$_{o}$ɑ	: r_s = .75	.005
	ˈi	: r_s = .96	.0005
	S$_{b}$	: r_s = .82	.025
	ˈɑ$_{min}$-ŋ$_{max}$	: r_s = .71	.025
	ˈɑ$^{2}_{3}$-ŋ$_{max}$	: r_s = .79	.025
	ˈu-$_{o}$ʌ	: r_s = .24	-
Cuts no. 2 and 3	ˈɑ$_{min}$	: r_s = .07	-
	ˈɑ$^{2}_{3}$	: r_s = -.08	-
	ŋ$_{max}$	: r_s = .63	.10
	ˈu	: r_s = .96	.0005
	$_{o}$ʌ	: r_s = .96	.0005
	$_{o}$ɑ (cut 2, only)	: r_s = .86	.01

(continued from the preceding page) <u>Table 11</u>

Correlation between response rankings and rankings of				level of significance
<u>Cuts no. 2 and 3</u> (continued)	$'\alpha_{min}-\eta_{max}$	:	r_s = .65	.01
	$'\alpha_{3}^{2}-\eta_{max}$	:	r_s = .75	.005
	$'u-{}_{o}\wedge$	:	r_s = .19	-
<u>Cut no. 4</u>	$'\alpha_{min}$	:	r_s = -.07	-
	$'\alpha_{3}^{2}$	:	r_s = .75	.005
	η_{max}	:	r_s = .58	-
	$'u$	:	r_s = .86	.01
	$'\alpha_{min}-\eta_{max}$	:	r_s = .71	.05
	$'\alpha_{3}^{2}-\eta_{max}$	:	r_s = .75	.005
	$'\alpha_{min}-'u$	:	r_s = .85	.01
	$'\alpha_{3}^{2}-'u$	:	r_s = .89	.005
<u>Cut no. 5</u>	$'\alpha_{min}$	:	r_s = .21	-
	$'\alpha_{3}^{2}$	:	r_s = .03	-
	$'\alpha_{min}-\eta_{max}$	:	r_s = .82	.025
	$'\alpha_{3}^{2}-\eta_{max}$	:	r_s = .79	.025
	η_{max}	:	r_s = .88	.005

4.1.3.3 Cut no. 3: "mange busser" - test 3

The top in the identification function for the non-final category is rather broad-banded, cf. fig. 10, mid right, but the shapes of the three curves still bear a good resemblance to fig. 10, top left. The order of stimuli is the same as in cut no. 2, i.e. NF-3 is shifted one to the left, compared to fig. 10, top left, and the proportion of interrogative-responses has risen at the expense of the non-final category, compared to the parent stimulus. The correlations between responses and Fo are the same as for cut no. 2.

4.1.3.4 Cut no. 4: "mange bus" - test 3

Even in this case, fig. 10 lower left, three peaks in the identification functions appear, and the order of stimuli is again only disturbed in one place, as compared to fig. 10, top left, i.e. NF-3 has been shifted one to the right. Compared to fig. 10 top left, the proportion of declarative-responses has risen, mainly at the expense of the interrogative-category. Compared to fig. 10, mid right (cut no. 3), both the declarative and non-final categories grow, whereas the interrogative one shrinks. This last pattern was also observed between parent stimuli and cut no. 1, and serves to confirm the explanation offered for this phenomenon in section 4.1.3.1.

In table 11, "cut no. 4", have been included the correlations between responses and the measured rise from [ˈɑ] to [ˈu], which are seen to be as high as (or higher than) that between responses and [ˈu]. (And, as in test 1, $ˈ\alpha_3^2$ still renders a slightly better correlation than $ˈ\alpha_{min}$.) This rise is indicative of the slope between the two stressed vowels, and seems to play the role here that S_b did in test 1 (and cut no. 1).

As long as two stressed vowels remain (which together suffice to determine a slope), identification is not very seriously affected - and is not very different from identification of the parent stimuli.

4.1.3.5 Cut no. 5: "mange" - test 3

Now the picture changes completely. The order of stimuli in fig. 10, lower right, is a complete mess, the groups no longer resemble those of fig. 10, top left, and only one stimulus, D-4, received unambiguous responses. The non-final responses dominate. The strongest correlation between responses and fundamental frequency is found in the level of η_{max}, cf. table 11, but since this point does not correlate too well with the rest of the contour, cf. table 6, there is no real cue in it as to the kind of utterance it occurred in.

4.1.4 Conclusion - test 3

It seems that the predictions made in the introduction to tests 3 and 4 (section 4.) are substantiated by the results. Earlier parts of the utterance do carry information about the intonation contour and identification deteriorates but slightly when stimuli are mutilated - until nothing is left but the first stress group, which contains no relevant information. Subjects use all the cues they are given, still, however, with a slight predominance of the level of the last stressed (and unstressed) syllable in the utterance. - And now for the second part of the prediction, that identification should not be affected much by the removal of the first and second stress groups:

4.2 Test 4

Test 4 comprised stimuli with the first and second stress groups cut away from the beginning.

4.2.1 Stimuli - test 4

9 of the 15 stimuli from test 1 served as parents for test 4. They were: I-1, I-2, I-3 - NF-1, NF-2, NF-3 - D-1, D-3, D-4, i.e. the three best identified stimuli from each category (disregarding D-5). For a discussion of this choice, see section 4.1.1.

By means of a segmentator, the 9 parents were deprived of the first and second stress groups, thus:

cut no. 1: busser fra Tiflis

cut no. 2: Tiflis

9 parents times 2 cuts give a total of 18 stimuli. A test tape containing 5x18 randomized stimuli, plus 6 introductory dummies, i.e. 96 stimuli in all, was prepared. Stimuli were again introduced by numbers (read by the author) with a 3 sec interval for responding. The total duration of one run of the test tape was 11 minutes.

Subjects were instructed exactly as in tests 1 and 3, cf. section 2.2.

The test tape was presented twice, with at least one day's interval, giving a total of 10 responses to each stimulus by each subject.

4.2.2 Subjects - test 4

The same subjects were employed as served in test 3. (In fact, they took test 3 and test 4 intermittently.)

4.2.3 Results - test 4

The results of test 4 may be got over with quickly: Fig. 11, top, and table 12 present the 4 subjects' distribution of responses in test 1 to the 9 parent stimuli of test 4. Compared to fig. 3, NF-1 and NF-3 have changed places, but the Spearman rank correlation coefficient between response rankings by 10 subjects in test 1 to the 9 stimuli of test 4 and response rankings by 4 subjects to these stimuli is .98 - and the 4 subjects are still, of course, good representatives of the original 10.

Fig. 11 and table 13 present the results. (In fig. 11, stimuli have been grouped as in the earlier figures, cf. sections 2.4 and 3.1.)

4.2.4 Cut no. 1: "busser fra Tiflis" - test 4

Fig. 11, mid, shows three very clear peaks in the identification functions. In fact, the categories are more clear-cut than those for the parent stimuli. The task must have been easy, probably due to the fact that there were fewer different stimuli than in test 1.

In table 14 are given Spearman rank correlations coefficients between rankings of responses, on the one hand, and rankings of a number of the Fo measurements presented in table 4.[1] [ₒi], [ˈi], [ˈu], S_t, [ₒʌ], and S_b all correlate highly with responses, [ₒi] slightly more so than the others, and the same conclusion as about test 1, section 2.4.2 may be made: that these parameters are all intercorrelated and no one single parameter can be made much more responsible for identification than the others.

4.2.5 Cut no. 2: "Tiflis" - test 4

Here again, fig. 11, bottom, subjects evidently found no difficulty in recognizing three categories, and in fact, the distribution of responses almost exactly is 1/3 to each category, cf. table 13, "cut 2". Compared to the parent stimuli, cut no. 2 received a higher proportion of declarative-responses and a smaller proportion of interrogative responses, i.e. a certain number of responses were moved one down on the interrogative-declarative scale (leaving the non-final category intact). The same sort of argument that was used in section 4.1.3.1 about the difference between parent stimuli and "mange busser fra Tif" may serve again: The original material for analysis also contained one-word statements and questions, among them "Tiflis." and "Tiflis?". Not only did the level of these words vary, but so did the Fo movement in the stressed syllable, which was falling

---

1) Note, again, that the coefficients in table 14 are not directly comparable to those in table 5, being based on rankings of only 9 items.

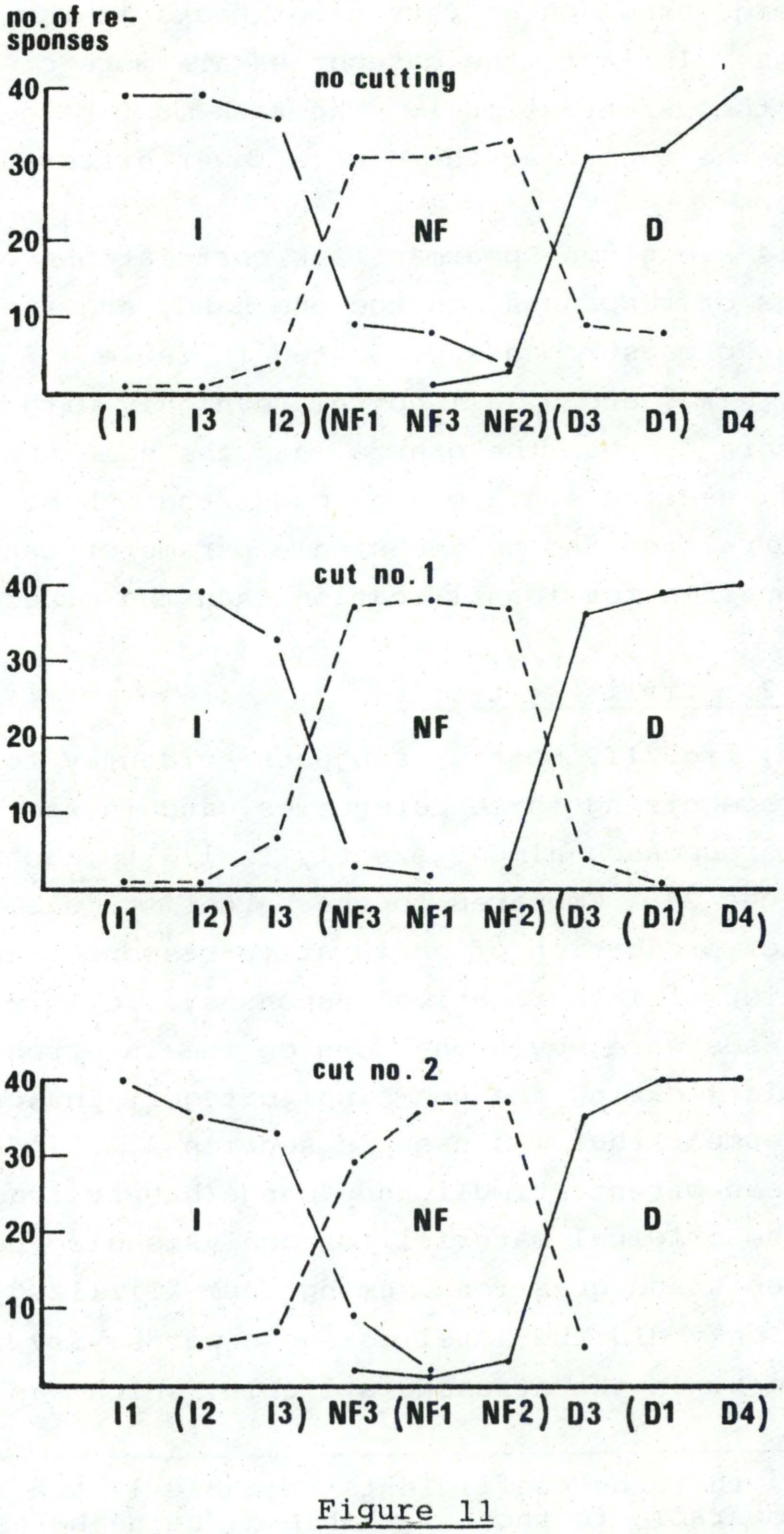


Figure 11

Identification functions for the 9 stimuli in test 4, as identified by 4 Copenhagen speakers.

Table 12

4 Copenhagen speakers' response in test 1 to the 9 stimuli of test 4.

response / stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
D-1	32	8		40
D-3	31	9		40
D-4	40			40
NF-1		31	9	40
NF-2	3	33	4	40
NF-3	1	31	8	40
I-1		1	39	40
I-2		4	36	40
I-3		1	39	40
Total	107	118	135	360

Table 13

4 Copenhagen speakers' response to the 9 stimuli of test 4.

response / stimulus	DECLARATIVE	NON-FINAL	INTERROGATIVE	Total
cut 1				
D-1	39	1		40
D-3	36	4		40
D-4	40			40
NF-1		38	2	40
NF-2	3	37		40
NF-3		37	3	40
I-1		1	39	40
I-2		1	39	40
I-3		7	33	40
Total	118	126	116	360
cut 2				
D-1	40			40
D-3	35	5		40
D-4	40			40
NF-1	1	37	2	40
NF-2	3	37		40
NF-3	2	29	9	40
I-1			40	40
I-2		5	35	40
I-3		7	33	40
Total	121	120	119	360

Table 14

Spearman rank correlation coefficients, r_s, between the rankings of responses to the stimuli of test 4, on one hand, and the rankings of certain Fo measurements in these stimuli, cf. table 4. The level of significance (if better than .10) is indicated in the right hand column.

Correlation between response rankings and rankings of				level of significance
Cut. no. 1	ˈu	:	$r_s = .95$	.0005
	˳ʌ	:	$r_s = .91$	.0005
	˳ɑ	:	$r_s = .80$	.0005
	ˈi	:	$r_s = .95$	.0005
	˳i	:	$r_s = .98$	.0005
	S_b	:	$r_s = .91$	.0005
	S_t	:	$r_s = .95$	.0005
	ˈu-˳ʌ	:	$r_s = .33$	-
	ˈi-˳i	:	$r_s = .66$	.05
Cut no. 2	ˈi	:	$r_s = .97$	.0005
	˳i	:	$r_s = .99$	.0005
	ˈi-˳i	:	$r_s = .68$	.025

in the statement and rising in the question. This variation was not found in the longer sentences. Apparently, in one word utterances Fo movement in the stressed vowel takes over and compensates for the lack of a contour proper. And this lack of compensation in the mutilated stimuli of test 4's cut no. 2 may account for the discrepancy between fig. 11 top and bottom.

As a matter of fact, it might be argued that since cut no. 2 received an almost perfect distribution of responses (very nearly 1/3 in each category), the lack of a compensatory Fo movement in the stressed vowel of "Tiflis" had no effect at all, and should therefore be deemed unimportant for the identification of intonation contours. (In this connection it is only fair to say that we do not know what the identification functions for the parent stimuli of tests 3 and 4 had looked like, had they been the only stimuli in test 1 - there might have been very little discrepancy to account for between identification functions in test 3, cut 1, and test 4, cut 2, respectively, and the parents.)

The correlation between responses and [̥i] and [ˈi] are very high indeed, in fact, it almost reaches unity with [̥i].

4.2.6 Conclusion - test 4

Evidently, and not surprisingly, the end of the utterance alone may suffice for identification of the intonation contour it rides upon. Thus, the predictions made in the introduction to tests 3 and 4 (section 4.) are borne out, and the model (fig. 1) as well as the results of test 1 are confirmed.

5. General conclusion and discussion

In a previous analysis of Advanced Standard Copenhagen Danish intonation, the intonation contour (as defined by the course of the stressed syllables in an utterance) was found to vary according to sentence type, being most steeply falling in declarative sentences, flat in statement-questions, and varying in between these two extremes in other types of questions and in non-final periods.

A certain trade-off between intonation contours and syntax[1] was found: the more syntactic information about the non-declarative function is contained in a sentence, the more declarative-like is its intonation contour. Thus, we may acoustically recognize roughly three types of intonation contours, but each with a certain dispersion and with a certain overlap as well. As it turned out, Copenhagen listeners can identify such utterances solely on the basis of their fundamental frequency course (i.e. given no contextual and no syntactical information): the less falling the intonation contour, the more interrogative utterances are heard, the more steeply falling the contour, the more declarative utterances are heard, and contours in the middle of the continuum are identified as being non-final. (Note that identification is not always in agreement with the speaker's intention, however.) Judging from the identification functions, perception seems to be categorical: three clear peaks are found, one for each category of interrogative, non-final, and declarative, respectively. (Non-Copenhagen speakers, on the other hand, did not, in these 15 stimuli by a Copenhagen speaker, identify three categories, but any conclusions as to the reason why must await an acoustical analysis of intonation contours in various dialects.)

Close inspection of the course of fundamental frequency in the stimuli revealed rather high correlations between response distributions, on one hand, and the course of Fo, on the other, with a slight predominance of the <u>level</u> of the final stress group in the utterance (whereas the final rise from stressed to unstressed syllable gave no such high correlation to responses). However, as all points of measurements during the second and third stress groups were highly correlated/concordant (ie., if one point is high, so are all the others), and since partial rank correlations could not determine one single parameter as being more independent of the others than any other parameter, we may conclude that identification is determined not just by the level of the final part of Fo, but by earlier parts as well. That this is actually the case was demonstrated by the identification

---

1) See the footnote on page 2o.

of stimuli where a greater or lesser number of syllables had been cut away from the end. Only when just the first stress group remains is identification seriously affected. This is not surprising, since intonation contours are least different in this first part of the utterance and since, further, the course of Fo during the first stress group showed a rather low correlation with later parts. - On the other hand, utterances are identified very well indeed when only the last stress group is presented, which is also to be expected, since this is the stretch where different contours are most widely differentiated acoustically. - All in all, we may conclude with Bolinger (1951, p. 206) that "... the basic entity of intonation is a pattern, not a pattern in the relatively abstract sense of grammatical recurrences, but in the fundamental, down-to-earth sense of a continuous line that can be traced on a piece of paper."

As for the number and kinds of categories established: Three perceptual categories of intonation contours seem to be the rule, rather than the exception: Gårding and Abramson (1965) established three in American English, "neutral statement", "yes or no question", and "counting in a series" (plus two that I would be inclined to call attitudinal: "anger", and "delighted surprise"). Hadding-Koch and Studdert-Kennedy (1963, 1965, 1974) found for Swedish and American subjects three categories, "statement", "question", and "talking-to-self" (the linguistic status of this last category is not transparent and was one whose meaning they had some difficulty in explaining to their subjects - I suspect that labelling it "non-final" or "continuation" might have made the task easier for the subjects and rendered the same results). Isačenko and Schädlich (1963) found three categories in German, "nicht-Frage", "progredient" or "weiterweisend", and "Frage". Johansson (1978) established three categories in Swedish and labelled them "declarative", "continuation", and "interrogative". Uldall (1961) found two categories in English, "question" and "statement" - but some stimuli were highly ambiguous and one is left to wonder whether the introduction of a third category might not have resolved this ambiguity.

The acoustical manifestation of the three categories varies from one language to another (although the similarities may be greater than the differences) and so do the perceptual cues to their identification (but even the perceptual processing of intonation contours may be less divergent than different authors on the subject would have us believe). Thus, any variation aside, there is little doubt that in a number of languages (or dialects) subjects are able to identify three types of intonation contours. Whether they all represent a linguistic reality is another question: it is characteristic that most of the perceptual experiments have had subjects work in a forced choice situation (Uldall, 1961, is an exception). Sven Öhman has objected (in a personal communication) that this is a doubtful procedure, and that the suspect third category, "non-final" or "continuation", may be a waste-paper basket into which everything is thrown that does not fit either the declarative or interrogative categories. - Had all the investigations been conducted with synthetic stimuli (and without prior acoustic analyses), one would have had to give the non-final category the benefit of a doubt. But the present investigation, and partly also those of Gårding and Abramson (1965) and Johansson (1978), has been carried out with real speech stimuli, chosen from a set of statements, questions,and non-final periods whose intonation contours varied accordingly, and subjects were thus not confronted with stimuli that are unlikely to occur in real speech, on the contrary: what they heard and identified was something that speakers actually produce, i.e. a linguistic reality. Therefore, I think, there are good grounds for accepting declarative, non-final, and interrogative intonation contours on an equal footing. (We may even have to continue the search, cf. Mikoš (1976) who found three types of interrogative contours, "yes-no questions", "wh-questions", and "alternative questions", plus a "declarative".)

Postscriptum

Although I do not, of course, wish to throw doubt upon the results reported above to a degree that would make them inconclusive and invalid, I do think it necessary to add a word of reservation: The four tests were based on only 15 utterances by a single speaker. This speaker, SH, is probably neither more, nor less, typical of ASC Danish than any of the other three speakers employed in the previous analysis, and the 15 stimuli were chosen among SH's recordings mainly for "technical" reasons, cf. section 2.1.1. We cannot know for sure (only make qualified guesses about) how the listeners would have reacted to other stimuli by SH or to stimuli by other ASC speakers. (And we do not know whether the non-ASC subjects might not have identified three categories of intonation contours in utterances by speakers of their various dialects.) Likewise, we do not know how great, or how small, a relative weight sentence intonation contours carry in speech perception in real-life communication where context and syntactic information are not suppressed the way they were in the present experiments.

References

Bolinger, D.L. 1951: "Intonation: levels versus configurations", Word 7, p. 199-210

Bolinger, D.L. 1958: "A theory of pitch accent in English", Word 14, p. 109-149

Bolinger, D.L. 1964: "Intonation as a universal", Proc. Ling. IX, 833-48

Bolinger, D.L. 1970: "Relative height" in: P. Léon, G. Faure, and A. Rigault (eds.) Prosodic Feature Analysis, Paris, p. 109-125

Breckenridge, J. and M.Y. Liberman 1978: "The declination effect in perception", Bell Laboratories (manus.)

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

Bruce, G. and E. Gårding 1978: "A prosodic typology for Swedish dialects", paper read at the Symposium on the prosody of Nordic languages, Lund 14-17 June (will appear in Travaux de l'Institut de Linguistique de Lund 13)

Cohen, A. and J. t'Hart 1967: "On the anatomy of intonation", Lingua 19, 177-192

Collier, R. and J. t'Hart 1975: "The role of intonation in speech perception" in: Cohen and Nooteboom (eds.) Structure and process in speech perception, Springer Verlag

Daneš, F. 1960: "Sentence intonation from a functional point of view", Word 16, 34-54

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

von Essen, O. 1956: "Hochdeutsche Satzmelodie", Zs.f.Ph. IX, 1, 75-85

Gårding, E. and A.S. Abramson 1965: "A study of the perception of some American English intonation contours", Studia Linguistica 19, 61-79

Hadding-Koch, K. 1961: Acoustic-phonetic studies in the intonation of southern Swedish, Gleerup, Lund

Hadding-Koch, K. and M. Studdert-Kennedy 1963: "A study of semantic and psychophysical test responses to controlled variations in fundamental frequency", Studia Linguistica 17, 65-76

Hadding-Koch, K. and M. Studdert-Kennedy 1964: "An experimental study of some intonation contours", Phonetica 11, 175-185

Hadding-Koch, K. and M. Studdert-Kennedy 1965: "Intonation contours evaluated by American and Swedish test subjects", Proc.Phon. V, 326-331

Hadding-Koch, K. and M. Studdert-Kennedy 1974: "Are you asking me, telling me, or talking to yourself?", J.Ph. 2, 1, 7-14

t'Hart, J. 1966: "Perceptual analysis of Dutch intonation features", IPO Annual Progress Report 1, 47-51

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

Isačenko, A.V. 1965: "Deutsche Satzintonation als Konfigurationen zweier diskreter Tonebenen (mit Demonstration von Tonbändern)", Proc.Phon. V, 371-373

Isačenko, A.V. and H.-J. Schädlich 1963: "Untersuchungen über die deutsche Satzintonation", Studia Grammatica 7, 7-67

Johansson, I. 1978: Funktionella aspekter på satsintonationen i svenska (Functional aspects of sentence intonation in Swedish), Acta Universitas Umensis, Umeå Studies in the Humanities 16

Lehiste, I. and G.E. Peterson 1961: "Some basic considerations in the analysis of intonation", JASA 33, p. 419-425

Mikoš, M.J. 1976: "Intonation of questions in Polish", J.Ph. 4, 247-253

Reinholt Petersen, N. 1976: "Intrinsic fundamental frequency of Danish vowels", ARIPUC 10, p. 1-27

Rossi, M. 1971: "Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole", Phonetica 23, p. 1-33

Rossi, M. 1978: "La perception des glissandos descendants dans les contours prosodiques", Phonetica 35, p. 11-40

Studdert-Kennedy, M. and K. Hadding 1972: "Further experimental studies of fundamental frequency contours", Proc.Phon. 7, 1024-1031

Studdert-Kennedy, M. and K. Hadding 1973: "Auditory and linguistic processes in the perception of intonation contours", L&S 16, 4, 293-313

Thorsen, N. 1976: "An acoustical investigation of Danish intonation: preliminary results", ARIPUC 10, p. 85-147 (will also appear, in slightly revised form, in J.Ph. 6, 1978)

Thorsen, N. 1977: "On the interpretation of raw fundamental frequency tracings", ARIPUC 11, p. 1-11

Thorsen, N. forthcoming: "Interpreting raw fundamental frequency tracings of Danish" (an enlarged version of Thorsen, 1977, to appear in Phonetica)

Uldall, E. 1960: "Attitudinal meanings conveyed by intonation contours", L&S 3, 223-234

Uldall, E. 1961: "Ambiguity: question or statement? or 'Are you asking me or telling me?'", Proc.Phon. 4, 779-783