Reprinted from NORDIC PROSODY. Gårding, Bruce, Bannert (eds). 1978

ASPECTS OF DANISH INTONATION

Nina Thorsen

A MODEL OF DANISH INTONATION

Intonation in short sentences in Advanced Standard Copenhagen Danish (ASC) may be accounted for in a model, cf. Figure 1. (For a detailed account of the procedure that led to the formulation of this model, see Thorsen, 1976.) A basic assumption, underlying the model, 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 Figure 1). (3) To the resultant of

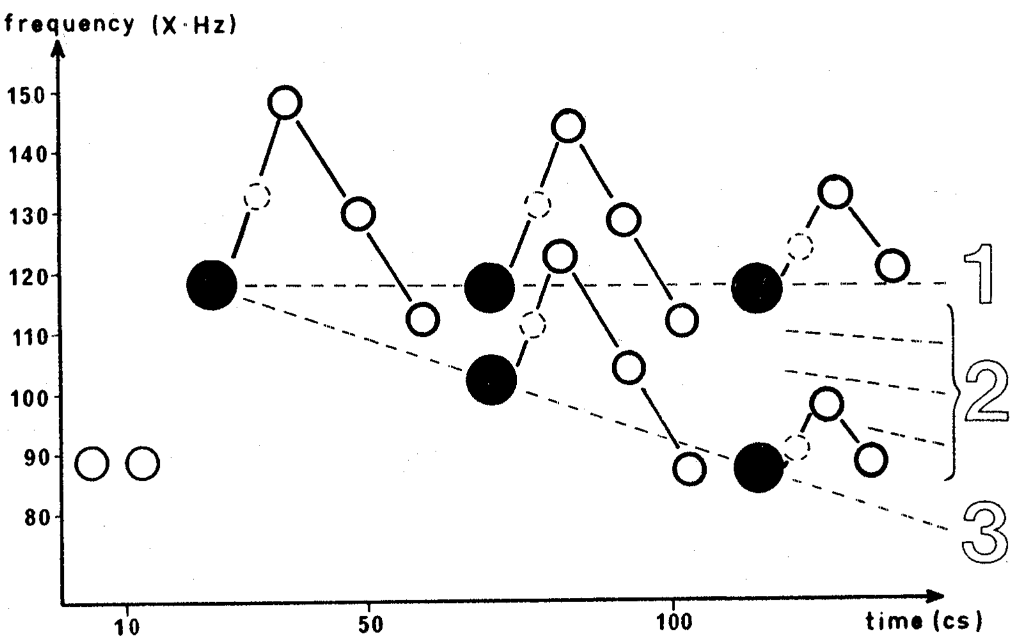

Figure 1. A model for fundamental frequency in short sentences in Advanced Standard Copenhagen Danish. 1: statement questions, 2: interrogative sentences with word order inversion and/or with interrogative particle, and non-final periods (variable), 3: declarative sentences. The heavy circles represent 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.

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 Fo level differences between 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 (1977 and forthcoming).

As may be seen from the full lines of figure 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 ASC 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 broken lines of figure 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).

Figure 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. A feature common to all three languages is the fact that both lines decline for statements (more or less steeply), but the top-line declines faster than the base-line, so that top- and base-lines together create a wedge-shape.

The intonation contour tends to vary systematically with sentence type, as suggested by figure 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 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, p. 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).

PERCEPTION OF INTONATION CONTOURS

The observation of the trade-off between syntax and intonation contour led to the formulation of the following questions:

(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 utterance are the various contours perceptually differentiated?

Space does not allow for a detailed account of the four tests that were designed to answer the above questions (the reader is referred to Thorsen 1978) but a summary of the results will be presented, as well as a discussion of "terminal rises".

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 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 was the fact that it has not been done with Danish material before. Furthermore, the material and procedure deviate somewhat from those of previous investigations.

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

Five recordings by a male subject of a statement Der går mange busser fra Tiflis. 'There are many buses out of Tiflis.', a 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 a question with word order inversion Går der mange busser fra Tiflis? 'Are there many buses out of Tiflis?' were selected for identification. By means of a segmentator the sequence ['maŋ̩ 'b̥usʌ fʁ̥o 'd̥ˢifl̥is] was isolated from the 15 items. The 15 stimuli constituted a quasi-continuum on the "slope-scale", ranging from flat to steeply falling contours. Durational differences between segments across the three sentence types were small and in no case statistically significant, so we may assume that subjects' judgments were based solely upon the Fo course in the stimuli.

Test 1

In test 1 (designed to answer the first question above) 10 ASC speaking subjects identified stimuli as being either declarative, non-final, or interrogative (forced choice). Three clear peaks in the identification functions appeared, pointing towards categorical perception. Further, responses are highly correlated with the course of Fo, i.e. the higher Fo runs, the more responses for interrogative, and the lower Fo runs, the more responses for declarative (irrespective of the speaker's intention): A number of Fo measurements were performed in the 15 stimuli and Spearman rank correlation coefficients between responses and each of these measures were calculated. It turned out that responses correlated very highly with Fo of the last stressed vowel, ['i] (r_s=.98), with the succeeding unstressed vowel, [i̥] (r_s=.98), and with the slope of the intonation contour (i.e. the base-line) (as defined by the Fo difference between ['u] and ['i], and their time interval) (r_s=.95). Correlations were high between responses and the level of Fo in the second stressed vowel, ['u] (r_s=.89), the succeeding unstressed vowel, [ˌʌ] (r_s=.77), and the top-line (as defined by the Fo difference between [ˌʌ] and [ˌi], and their time interval) (r_s=.88). Correlations were rather low, however, between responses and Fo in the first stress group (['ɑ]:r_s=.18, [ŋ̍]:r_s=.64) and also low between responses and the rise from stressed to unstressed vowel in the second and third stress groups (['u]-[ˌʌ]:r_s=.38, ['i]-[ˌi]:r_s=.30). — The rather high correlations between responses and all of the Fo course in the second and third stress groups and with top-line and base-line slopes are, of course, partly due to the fact that these parameters are interdependent: Spearman rank correlation coefficients between a number of pairs of Fo measurements, as well as the Kendall coefficient of concordance between several such measurements were calculated, and it turned out that during the second and third stress groups all points correlate highly with all other points and with top-line and base-line slopes as well - the correlation or concordance being the higher, the closer the points of measurement are in time. In other words, if one point (or top-line/ base-line) is high, so are all other points of the Fo course, and vice versa. Further, Kendall partial rank correlation coefficients could not single out any one parameter which was more independent of the others than any other parameter.

In conclusion, we may say about the results of test 1 that they confirm the implication of the model in figure 1: intonation contours are more widely separated in later than in early parts of the utterance. Since the correlation/concordance between points in the Fo course in the second and third stress groups are high, 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, cf. the high correlation coefficients between responses and ['i] and [ˌi], respectively.

Terminal rises

It is interesting to note that the magnitude of the final rise, i.e. ['i]-[ˌi], shows only a slight (r_s=.30) 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, i.e. the levels of the final stressed and succeeding unstressed vowels, were.

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 resynthesized with the final unstressed syllable, -lis, being raised 60 Hz, in steps of 10 Hz (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 +30 and +40 Hz listeners reported that they heard a statement, but a rather obstinate one. At +50 Hz the utterance began to sound interrogative (to some listeners), and at +60 Hz 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 apply 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 is heard (in the stressed vowel of 'November'), 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.

Test 2

If the categorization performed by subjects in test 1 were purely phonetic, i.e. the continuum had been divided into three equally large bits, without any regard to linguistic analysis, then one would expect listeners to divide the same continuum into two equally sized bits when presented with only two categories to choose from. Accordingly, the test tape from test 1 was presented to 7 of the 10 subjects for identification as "declarative" or "non-declarative" (forced choice). (Test 2 was run 6 months after test 1 and subjects did know that the tapes were identical.) It turned out that the stimuli that were labelled "non-final" in test 1 were <u>not</u> split up into two equally sized portions, each one 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.

Tests 3 and 4

From the results of test 1, we might expect stimuli to be rather well identified even when the third stress group is missing, given that responses in test 1 correlated fairly highly with Fo of ['u] 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.

In test 3, 4 of the 10 ASC speakers were presented with mutilations of the 7 stimuli that had been identified best in test 1. One, two, three, four, and five syllables, respectively, were cut away from the end of these 7 utterances. The resulting 35 stimuli were presented, randomized, together. Subjects had to determine whether the cuts they heard originated from a declarative, a non-final, or an interrogative utterance (forced choice). It appeared that identification deteriorated but slightly, unless only the first stress group (<u>mange</u>) remained.

Test 4 comprised mutilations of the 9 stimuli that had been identified best in test 1. The first and the second stress groups were cut away. This has no apparent effect on identification, and we may conclude that although earlier parts of the Fo course do carry information about the intonation contour, the end of the utterance alone may suffice to identify the contour which it rides upon.

CONCLUSION

Thus, the predictions made above are borne out, and the model (fig. 1) as well as the results of test 1 are confirmed.

For a further discussion of the linguistic significance of these results, the reader is referred to Thorsen (1978).

REFERENCES

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

Bolinger D.L. 1964. Intonation as a universal. Proceedings of the Ninth International Congress of Linguists, 833-848

Bolinger D.L. 1970. Relative height. Prosodic Feature Analysis. Eds. P. Léon, G. Faure and A. Rigault. Paris, 109-125

Breckenridge J. and M.Y. Liberman. 1978. The declination effect in perception. Bell Laboratories, 38 pp. Manus.

Bruce G. 1977. Swedish word accents in sentence perspective. Travaux de l'Institut de Linguistique de Lund XII

Bruce G. and E. Gårding. 1978. A prosodic typology for Swedish dialects. Travaux de l'Institut de Linguistique de Lund XIII, 219-228

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. Structure and Process in Speech Perception. Eds. Cohen and Nooteboom. Berlin, 107-121

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. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 9, 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. Acoustico-phonetic studies in the intonation of Southern Swedish. Travaux de l'Institut de Phonétique de Lund III

Hadding-Koch K. and M. Studdert-Kennedy. 1963. A study of semantic and psycho-physical 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. Proceedings of the Fifth International Congress of Phonetic Sciences, 326-331

Hadding-Koch K. and M. Studdert-Kennedy. 1974. Are you asking me, telling me, or talking to yourself?. Journal of Phonetics 2, 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. Journal of Phonetics 1, 309-327

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 Universitatis Umensis, Umeå Studies in the Humanities 16

Lehiste I. and G.E. Peterson. 1961. Some basic considerations in the analysis of intonation. Journal of the Acoustical Society of America 33, 419-425

Mikoš M.J. 1976. Intonation of questions in Polish. Journal of Phonetics 4, 247-253

Studdert-Kennedy M. and K. Hadding. 1972. Further experimental studies of fundamental frequency contours. Proceedings of the Seventh International Congress of Phonetic Sciences, 1024-1031

Studdert-Kennedy M. and K. Hadding. 1973. Auditory and linguistic processes in the perception of intonation contours. Language and Speech 16, 293-313

Thorsen N. 1976. An acoustical investigation of Danish intonation: preliminary results. Annual Report of the Institute of Phonetics, University of Copenhagen 10, 85-147. (Also in Journal of Phonetics 6, 151-175, 1978)

Thorsen N. 1977. On the interpretation of raw fundamental frequency tracings. Annual Report of the Institute of Phonetics, University of Copenhagen 11, 1-11

Thorsen N. Interpreting raw fundamental frequency tracings of Danish. Phonetica. Forthcoming (an enlarged version of Thorsen, 1977)

Thorsen N. 1978. On the identification of selected Danish intonation contours. Annual Report of the Institute of Phonetics, University of Copenhagen 12, 17-73

Uldall E. 1960. Attitudinal meanings conveyed by intonation contours. Language and Speech 3, 223-234

Uldall E. 1961. Ambiguity: question or statement? or 'Are you asking me or telling me?'Proceedings of the Fourth International Congress of Phonetic Sciences, 779-783