PERCEPTUAL EXPERIMENTS WITH SWEDISH DISYLLABIC ACCENT-1 AND ACCENT-2 WORDS

Kurt Johansson

The Swedish - and Norwegian - so-called word tones have for a long time attracted a considerable amount of attention. According to Elert (1970) one may almost speak of something like a fixation to the problem, one reason for this being that the tone accent and its interpretation occupies a central place in the language system, and the way of solving the problem is essential for the interpretation of other prosodic elements as well.

Structurally, there exists an opposition between accent 1 and accent 2 (acute and grave accent, respectively) in words like ánden (the duck) and àndén (the spirit), stégen (the steps) and stègen (the ladder).[1] The opposition is only manifested in a stressed position.

There have been differences of opinion concerning the most relevant acoustic counterpart to the opposition. In most cases the choice has been between variations in the fundamental frequency (f_0) and the intensity, but also other differences, e.g. in duration and quality have been discussed.[2]

In 1955 Malmberg carried out a series of experiments with synthetic speech where the fundamental frequency, the intensity, and the duration were varied. Only f_0 was alone able to influence the judgments concerning minimal pairs like bûren (the cage) and bùren (carried).[3] This supported earlier supposi-

---

1 Elert (op.cit., p. 42 ff.) makes a distinction between this distinctive function and the connective function, where all syllables within the domain of the grave accent belong to one and the same word.

2 Malmberg (1962) mentions another possibility: syllable boundary.

3 The experiments have been reported in e.g. 1955, 1962, and 1967.

tions based on instrumental data, where the differences between accent-1 and accent-2 words had been manifested most consistently in the fundamental frequency (cf. fig. 1).

Acoustic investigations show that in most Swedish dialects there is a distinct difference between the forms of the frequency curves for accent-1 and accent-2 words, the remarkable thing being, however, that the patterns may be reversed if different dialects are compared (cf. fig. 2).[1] This was one of the reasons why Gjerdman (1954) expressed doubts concerning the distinctive function of the tone[2] and evidently considered the intensity[2] as relevant instead.[3] He suggested that either Malmberg or himself and Meyer (where he seems to have found support) had made some sort of mistake in their tonal investigations.

In the last few years there seems, however, to be considerable agreement on the tonal character of accent 1 and accent 2.

The present investigations

Results obtained with synthetic speech should, where possible, be verified with ordinary speech, but so far rather few perceptual experiments have been carried out. Further, the students on our phonetics courses had for a number of years, expressed their difficulty in identifying accent-1 and accent-2 words in different Swedish dialects. I therefore decided to prepare some tests.[4]

---

1 An interesting attempt to explain the change in the tone curves from one dialect to another has recently been made by Öhman (1967).

2 The distinction between different levels (i.e. perceptual and acoustic) is not consistent.

3 To support his view he states that he himself does not always make a tonal distinction between accent-1 and accent-2 words, and this he has found also with other speakers (p. 139).

4 The tests were carried out in spring 1967.

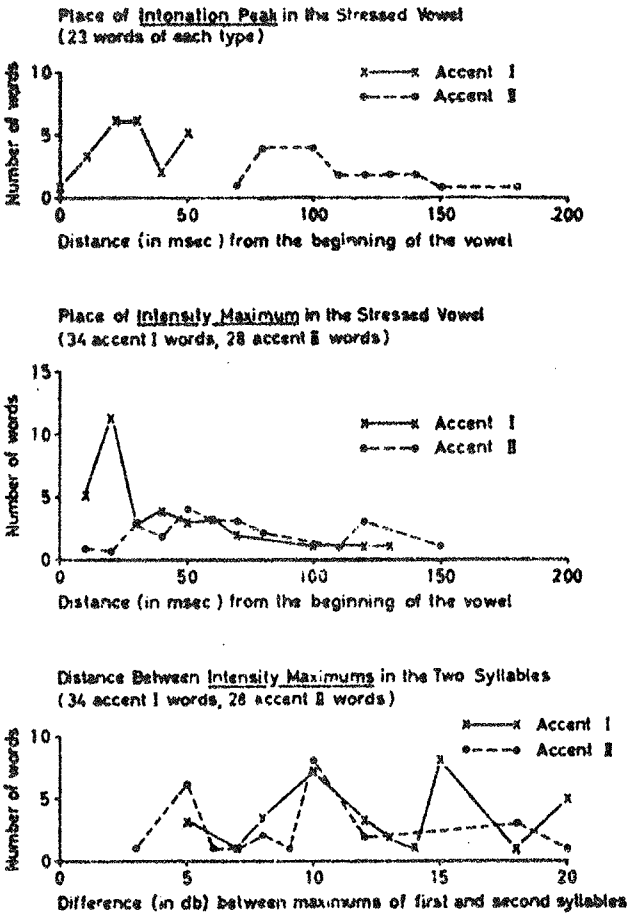

Fig. 1. Data on frequency and intensity overlapping in Swedish accent-1 and accent-2 words (Malmberg, 1962).

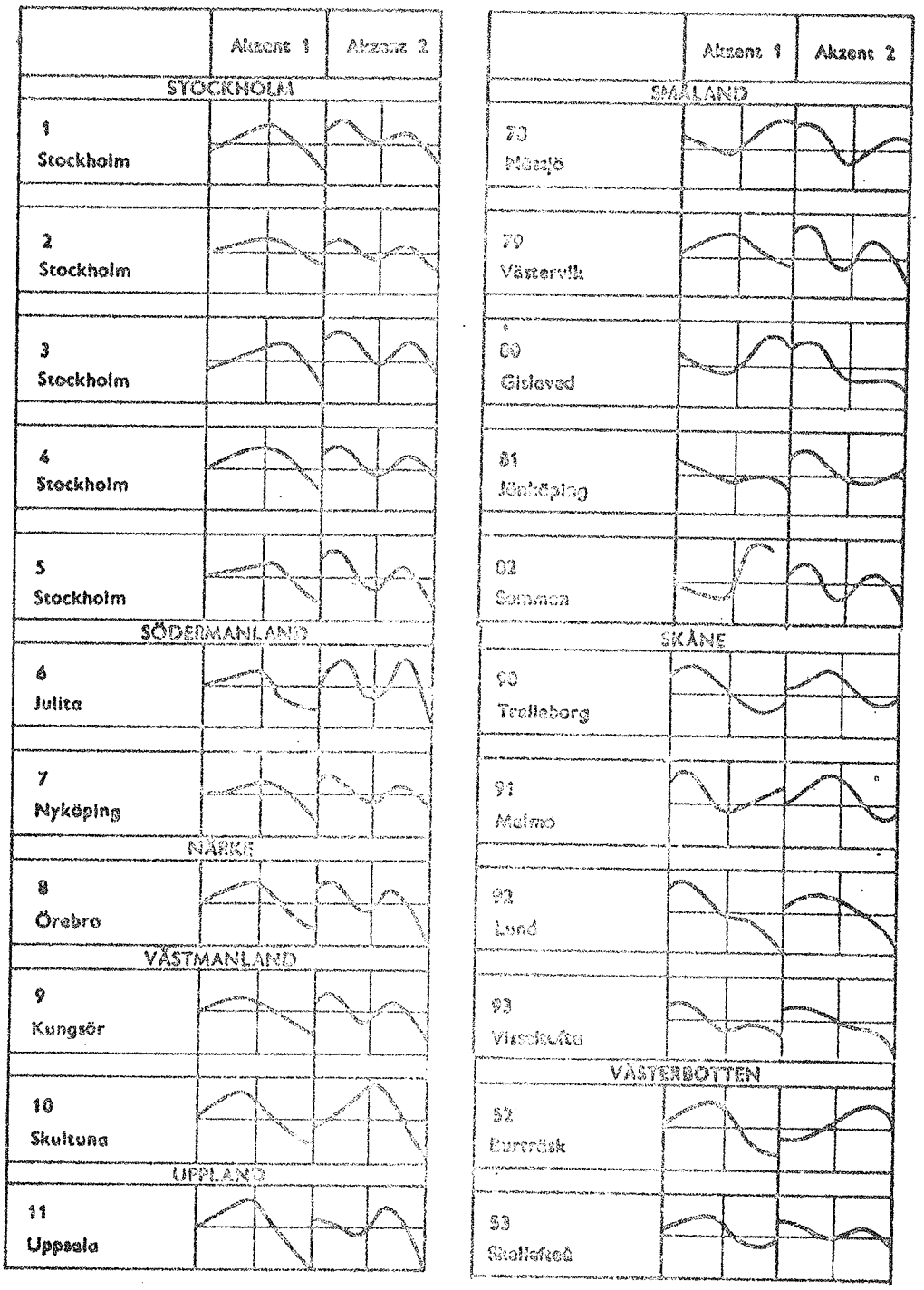

Fig. 2. Some frequency curves from Meyer (1954).

The dialect of speaker A, who had lived in or near Lund all his life, was typically Scanian[1] and the listeners had had good opportunities to become acquainted with his way of speaking (and also with the Scanian dialect itself since they were all studying at Lund University).

Speaker B was unknown to them, and one may suppose that this to a very high degree also goes for his West Bothnian (Umeå) dialect.[2] No listener spoke this dialect.

The f_0 curves of the first syllables, which are generally considered to contain the most important information, were virtually reversed for the speakers (cf. figs 3-4).

The following minimal pairs were recorded:

1) vå̂ken (the hole in the ice) : vȁken (awake)

2) skó̂tten (the shots) : skȍtten (the Scot)

3) tô̂mten (the [building] grounds) : tȍmten (the gnome)

The stressed syllables of the test words evidently contain the following segments in which the tonal differences might be manifested:

1) voiced contoid[3] + long vocoid

2) short vocoid

3) short vocoid + voiced contoid

The recordings were made on an Ampex 354 tape recorder (tape speed: 15"/sec) via a Neumann KM 54 condenser microphone. The words were repeated three times in succession, and the second examples were chosen as test words.

The test words were then transferred to a Lyrec TR_2 tape recorder speci-

---

1 Scania (Skåne), a province in the south of Sweden.

2 West Bothnia (Västerbotten) in the northeastern part of Sweden.

3 "The curve shape in the initial consonant, if there is any, may therefore be irrelevant to the distinction, as it has to be when the initial consonant is unvoiced." Fintoft, 1965.

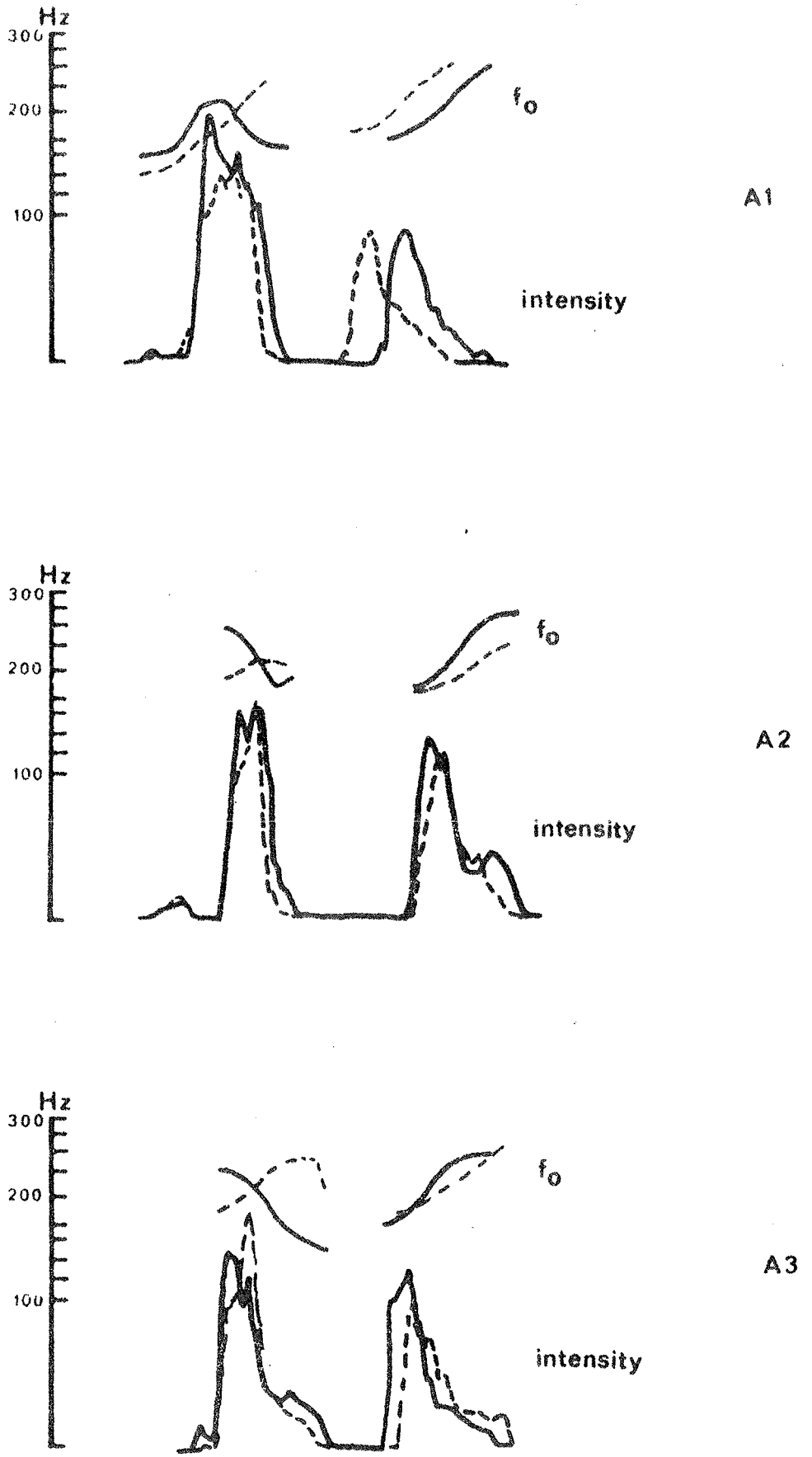

Fig. 3. Fo and intensity curves for speaker A's <u>våken</u>:<u>vȁken</u> (A1), <u>skȏtten</u>:<u>skȍtten</u> (A2), and <u>tȏmten</u>:<u>tȍmten</u> (A3). Broken lines = grave words.

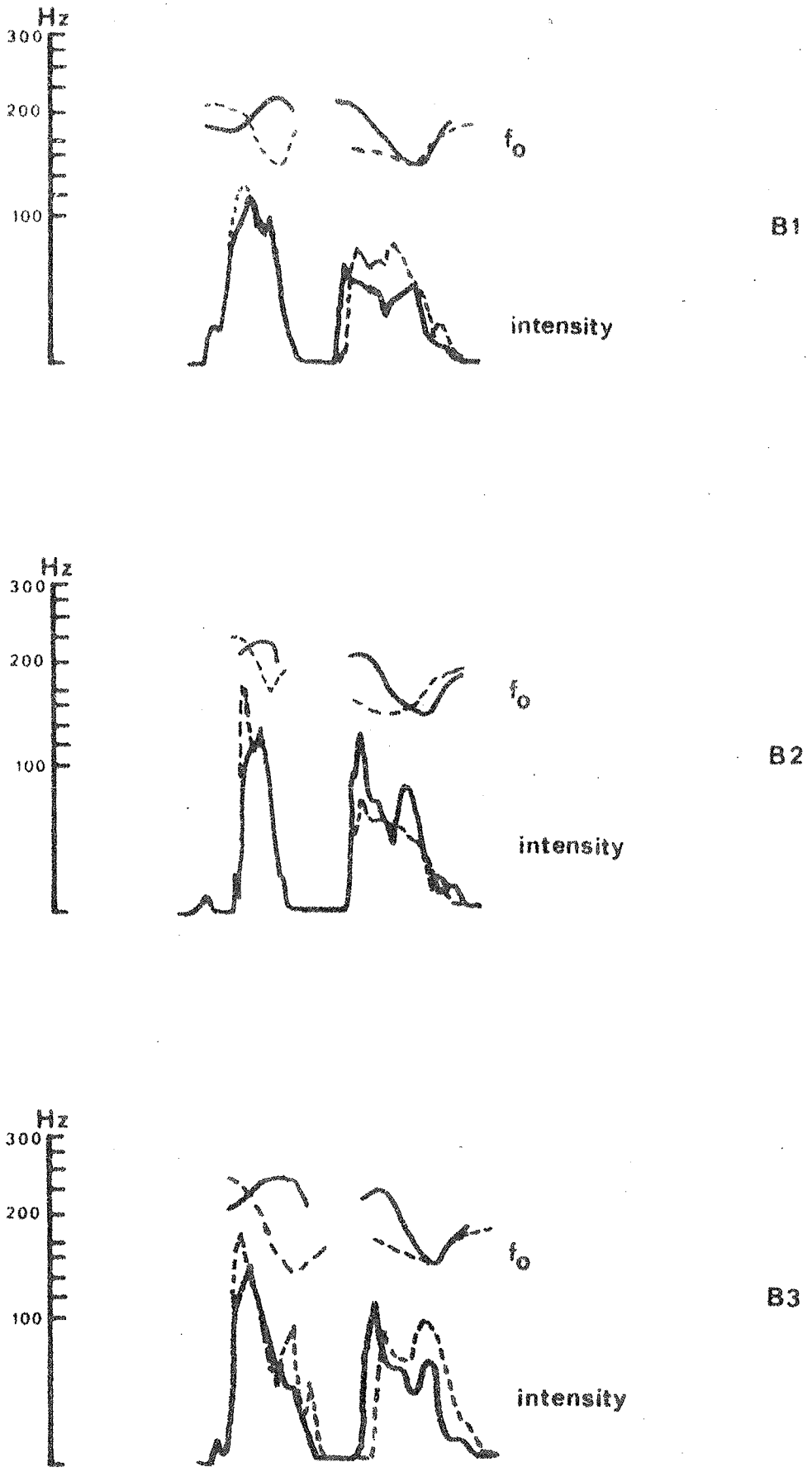

Fig. 4. Fo and intensity curves for speaker B's vå̀ken:vàken (B1), skɔ̀tten:skɔ̀ttən (B2), and tɔ̀mten:tòmten (B3). Broken lines = grave words.

ally adapted for use with an electronic gate. This tape recorder was also used when the cuts were made manually.

As a first step a loop was made for each test word. Then these words, with or without cuts, were transferred to a Revox F 36 tape recorder (tape speed: 7.5"/sec), which was also used when the stimuli were presented to the listeners (via a high quality loudspeaker[1]).

At the listening sessions the stimuli were presented in random order[2], the A-stimuli first in a separate group, then the B-stimuli. The only information given before the transition to the B-stimuli was that a different speaker would be heard.

Exactly the same stimulus was repeated three times in succession with one-second pauses, and this stimulus group then appeard five times in the test. The interval between different stimuli was about five seconds. After every tenth stimulus a longer pause was made (roughly ten seconds long). The first ten stimuli were presented, without the listeners' knowing it, only to acquaint them with the test procedure.

10-15 listeners took part at each listening session, which (including pauses) lasted about half an hour. The test words were presented in numbered frames on a black-board[3], and the listeners were instructed to mark each stimulus word "1" or "2" (forced choice).

Listeners

59 first-term students of phonetics were used as listeners. None of them had taken part in experiments of this type earlier.

1 Built at the Royal Institute of Technology, Stockholm.

2 Without any frame sentence.

3 Frames containing acute words were numbered 1, and grave words 2.

To be able to group the listeners according to "dialect", recordings were made of their pronunciation of the test words, in the same way as has been done with speakers A and B.

A rough estimation of the fundamental frequency curves, produced on a mingograph via a Trans Pitchmeter, then followed. The curves turned out to represent three main types with regard to the change in fundamental frequency in the stressed vocoid:

1. acute: falling grave: rising (or level)
2. acute: falling grave: falling (or level)

   In spite of the fact that both the acute and the grave patterns are mainly falling, there is a pronounced frequency difference between the two types, with accent 2 as the higher type.
3. acute: rising (or level) grave: falling

The first and largest group, mainly consisting of people from Scania, was subdivided in:

a. Native Scanians

b. Non-Scanians with f_0 patterns resembling the patterns for the Scanian acute and grave accents.

The following groups were established:

Group 1 a: 28 listeners from different parts of Scania.

Group 1 b: 9 listeners from the parts of Sweden nearest to Scania (2 from southern Halland, 6 from the southern parts of Småland, and 1 from Blekinge).

Group 2: 16 listeners chiefly from areas north of and nearest to the 1 b area (8 from Småland, 1 from Blekinge, 1 from Öland, 1 from Bohuslän, 2 from Västergötland, and 3 from Östergötland.

Group 3: 5 listeners from other districts (2 from the Stockholm area,

1 from Dalarna (Dalecarlia), 1 from Västmanland, and 1 from Ångermanland).

It should be observed that the boundary between groups 1 b and 2 is somewhat indistinct.

The curves for one of the listeners showed such irregularities that his score is only included in the first calculations under test I, where all listeners were treated as one group.

A comparison with the corresponding dialects in Meyer's and Malmberg's material, in so far as the dialects are represented there, seems in most cases to confirm the above group formation.

When studying the tables below it should be remembered that the groups, especially in some cases, are much too small to be considered representative of the dialect type in question, but as the results in spite of this may be of interest they are given for the sake of comparison.

Test I

This experiment was carried out in order to test to what extent listeners are dependent on their own dialect pattern for the identification of isolated acute and grave words.

In all, 12 stimuli were presented 5 times to 59 listeners, which gave 3540 judgments, or 295 per stimulus.

Results

A glance at the scores for the 59 listeners treated as one group (fig. 5) immediately reveals that Malmberg was right in his criticism of Gjerdman. The listeners, most of whom are Scanians, or at least from southern Sweden,

are much more successful in identifying the accent-1 and accent-2 words of speaker A. They are evidently to a large extent dependent on their own pattern for the identification.

	A	B
vâken	94 %	36 %
vàken	98 %	59 %
skôtten	97 %	47 %
skòtten	98 %	64 %
tômten	96 %	60 %
tòmten	95 %	67 %

Fig. 5. Percentages of correct responses to Scanian (A) and West Bothnian (B) stimuli (all listeners treated as one group).

Further, it can be seen from fig. 5 that there is a tendency to higher scores for the grave words, particularly the ones produced by speaker B.[1] The scores for this speaker's skòtten and tòmten deviate by +3σ from the means, for vâken by +2σ. They are obviously fairly sure that grave words had been produced. This is also the case for tômten (+3σ), while skôtten produces guessing, and as for vâken they give a score indicating that they interpret the word as grave instead of acute (-3σ).

In fig. 6 the number of correct responses is given for the different "dialect" groups. As these groups are of various sizes, resulting in variations between 150 and 840 judgments, the scores are given in percent for the sake of comparison.

---

1 In the following the score deviation from the theoretical means is given as ±nσ. E.g. ±3σ means that the deviation is at least 3σ, and that the listeners have with a high degree of certainty correctly identified the stimulus, while -3σ indicates that they are certain that another type of stimulus has been presented. ±2σ indicates relative certainty.

Group 1 a		Group 1 b		Group 2		Group 3	
A	B	A	B	A	B	A	B
99 %	52 %	99 %	50 %	99 %	52 %	68 %	92 %

Fig. 6. Percentages of correct responses to Scanian (A) and West Bothnian (B) stimuli, when the listeners have been divided into "dialect" groups.[1]

The first three groups show apparent similarities in their judgments. This is not very surprising as the geographical distribution is rather small, which should mean greater intercommunication possibilities and consequently greater familiarity with the respective dialects.

Group 2 had either only the acute f_0 pattern or the grave in common with the two speakers A and B. The reaction of this group to the stimuli suggests that groups 1 a, 1 b, and 2 all might have one common underlying perceptual pattern. But from this experiment alone, it is impossible to make any definite statement on this point (cf. test V).

Group 3 reacts in a quite different manner. The scores for both speakers deviate by +3σ, but there are considerably fewer correct responses for speaker A. (Groups 1–2 produced values indicating guessing for speaker B.) It is, however, probable that the rich opportunities for this group to become acquainted with the Scanian dialect (as well as with the speaker) have played an important role.

All groups gave considerably more correct responses to the grave words which might indicate that there is something in Hadding-Koch's tentative supposition that accent 2 for different dialects has some sort of common characteristic (1961, p. 69). The grave accent is, however, generally considered to be marked, and in a number of perceptual experiments I have no-

---

1 The groups included 28, 9, 16, and 5 listeners, respectively.

ticed that listeners tend to favour the marked member of an opposition, when a stimulus is felt to be somewhat "strange". This must also be taken into consideration.

The following conclusions could obviously be drawn: a) There are at least two perceptual patterns for the Swedish word tones. b) One's own dialect pattern is generally of vital importance for the identification of isolated accent-1 and accent-2 words. This is certainly demonstrated by the reactions of the Scanian listeners (group 1 a), where most Scanian stimuli are correctly identified, while West Bothnian stimuli produce guessing. c) Parenthetically it could be said that the responses to speaker B actually seem to be too high, if only the shape of the frequency curve of the first syllable is to be considered. One might in such a case have expected the correct responses to be very few, i.e. only a few percent. Some feature or other, besides the fundamental frequency of the first syllable, evidently contributes to the interpretation. There is of course more than one possibility, but one of them might be the obvious fundamental frequency differences in the second syllables for speaker B (cf. fig. 4). d) In any case it is evident that a few listeners have, at least to a certain extent, acquired a faculty to manage (two) different systems. A few listeners (only 4 in all) in groups 1 a, 1 b, and 2 have all 30 responses correct for both sets of stimuli and quite a few have almost as high a score for the "unfamiliar" dialect as for the dialect most resembling their own with regard to the f_0 pattern (in most cases however with the highest score for the latter type).

Test II

"I am inclined to regard the place of the peak within the stressed vowel

and the direction of the melody curve (mainly falling or rising) which follows from this difference, as the distinctive feature for word accent opposition." (Malmberg, 1955 and 1962). In the latter of these works he proposes, on a higher level of abstraction and "without having made the final auditory and synthetic tests", to denote the falling and rising curves "low" and "high", respectively.

There is no doubt that most authors agree with Malmberg that the fundamental frequency in the stressed syllable is the most relevant factor (see e.g. Fintoft, op.cit.).

The experiment below was intended to test the information in the isolated stressed syllables. With the aid of the electronic gate the last syllables were removed and the resulting stimuli were arranged as above.

In all, 12 stimuli were presented 5 times to 30 listeners, which gave 1800 judgments (900 per speaker).

Results

Fig. 7 gives the scores for the listeners treated as a single group.

	A	B
våken	83 %	19 %
vàken	83 %	21 %
skótten	93 %	31 %
skòtten	73 %	13 %
tómten	87 %	21 %
tòmten	79 %	18 %

Fig. 7. Percentages of correct responses to Scanian (A) and West Bothnian (B) stimuli (all listeners treated as one group).

The corresponding values for test I were considerably higher,[1] but for speaker A we still have high scores, while the scores for speaker B are very much lower than in test I (between 16 and 51 % lower for the different syllables).

All Scanian stimuli give scores that deviate by $+3\sigma$, the West Bothnian by -3σ. In the latter case the listeners are evidently sure that the acute words are grave and vice versa.

As the largest group of listeners belong to the Scanian or at least to the "South-Swedish" type, it seems from the lower scores in this test reasonable to assume that the second syllables possess a certain "distinctiveness", at least for the interpretation of some contour types.

This is confirmed by fig. 8, where a division into "dialect" groups has been made:

Group 1 a		Group 1 b		Group 2		Group 3	
A	B	A	B	A	B	A	B
85 %	14 %	91 %	18 %	87 %	25 %	32 %	53 %

Fig. 8. Percentages of correct responses to Scanian (A) and West Bothnian (B) stimuli, when the listeners have been divided into "dialect" groups.[2]

Considering the fact that some groups are very small a careful interpretation is needed. The first three groups show, however, once more evident similarities. Group 3 with only 2 listeners (from Ramsele and Stockholm) gives low scores also for speaker B, which might indicate that the second syllable is more important here. (Cf. test III.)

There is also some evidence that the second syllable might be of greater importance for the grave words. Cf. e.g. Hadding-Koch, op.cit., p. 66, where accent 2 is said to need its second syllable to be complete.

---

1 The differences are of about the same magnitude also if only the listeners of test II had been included in that test.

2 The groups included 15, 5, 8, and 2 listeners, respectively.

Test III

Test II suggested that the second syllables contain cues concerning the acute : grave distinction. In order to find out how strong these cues are, when the syllables are presented in isolation, the first syllables were eliminated with the electronic gate.

The fact that the listeners in this way heard twice as many -ten as -ken was not considered to matter very much.

In all, 12 stimuli were presented 5 times to 30 listeners, which gave 1800 judgments (i.e. 900 per speaker).

Results

Fig. 9 shows the scores when the listeners have been treated as a single group.

	A	B
våken	57 %	65 %
vàken	54 %	69 %
skôtten	65 %	73 %
skòtten	69 %	65 %
tômten	62 %	69 %
tòmten	47 %	71 %

Fig. 9. Percentages of correct responses to Scanian (A) and West Bothnian (C) stimuli (all listeners treated as one group).

In the following cases the number of correct responses deviate by +3σ: for all West Bothnian stimuli, and for Scanian skôtten and skòtten. Also tômten for this dialect indicates (relative) certainty (+2σ). In all other cases the listeners are guessing.

If we split into "dialect" groups, the results are as follows (fig. 10):

Group 1 a		Group 1 b		Group 2		Group 3	
A	B	A	B	A	B	A	B
60 %	67 %	61 %	66 %	59 %	69 %	52 %	92 %

Fig. 10. Percentages of correct responses to Scanian (A) and West Bothnian (B) stimuli, when the listeners have been divided into "dialect" groups.[1]

The scores for groups 1a and 1b deviate by +3σ (both speakers). For the West Bothnian stimuli, groups 2 and 3 have the same deviation, the score being considerably higher for group 3. The Scanian stimuli give for group 2 a deviation of +2σ (i.e. relative certainty), while we have guessing for group 3.

A remarkable thing is that also the groups 1a, 1b, and 2 have higher percentages for West Bothnian than for Scanian stimuli, and that these values are appreciably higher than the values for the corresponding first syllables in test II. Group 3 had better success for both speakers when listening to the second syllables and in fact gave for speaker B the same score as when the whole words were presented in test I.

Test IV

The second syllables evidently contain so much information that only one group for one of the speakers produces guessing. The investigation below was carried through in order to test how much this second-syllable information is able to influence the perception, if the first syllables from the acute words are combined with the second syllables of the grave words, and vice versa. The cuts were made manually just in front of the bursts, which are easy to locate auditively as well as instrumentally.

1 The groups included 15, 5, 8, and 2 listeners, respectively.

In all, 12 stimuli were presented 5 times to 29 listeners, which gave 1740 judgments (i.e. 870 per speaker).

Results

Fig. 11 shows the responses for the listeners treated as one group.

	A	B
vå-kèn	93 %	12 %
và-kén	96 %	5 %
skó-ttèn	96 %	10 %
skò-ttén	99 %	7 %
tóm-tèn	96 %	13 %
tòm-tén	98 %	10 %

Fig. 11. Percentages of responses (in accordance with the first syllables) to Scanian (A) and West Bothnian (B) stimuli (all listeners treated as one group).

The scores for the Scanian stimuli (reported in relation to the first syllables) are apparently very much the same as in test I where the unchanged test words had been presented, but higher than in test II.[1] For the West Bothnian stimuli they are much lower, generally considerably lower than in test II, where only the first syllables had been presented, and already those data were very low.

The second syllables evidently seem to be of greater importance, if the listeners (most of which are Scanians) hear the more unfamiliar dialect.

If we split into "dialect" groups the results are as in fig. 12.

---

1 The higher scores in this test as compared to test II probably depend on the more "natural" disyllabic stimuli here.

Group 1 a		Group 1 b		Group 2		Group 3	
A	B	A	B	A	B	A	B
99 %	2 %	100 %	16 %	100 %	14 %	55 %	38 %

Fig. 12. Percentages of responses (in accordance with the first syllables) to Scanian (A) and West Bothnian (B) stimuli, when the listeners have been divided into "dialect"groups.[1]

Also group 3 (consisting of only 2 listeners, however,) has difficulties in interpreting the West Bothnian stimuli. Judging from the scores for all groups the second syllables produced by speaker B contain more information than those produced by speaker A. This is what could be expected from the fundamental frequency curves (figs. 3-4). If, however, the greater fundamental frequency differences between the second syllables of the acute and the grave words for speaker B are symptomatic for most Central and Northern Swedish dialects as compared to the Southern type(s) is hard to say, but e.g. Meyer's data indicate such a difference.

Test V

In this test only the first syllables of våken and vàken were presented. In these syllables successive final cuts had been made with the electronic gate. At every cut portions of about 20 msec. were eliminated. Part of the investigation seems to be very much the same as has been done for Norwegian by Efremova et al. (1963).

In all, 42 stimuli were presented 5 times to 28 listeners, which resulted in 5880 judgments (3080 for speaker A, and 2800 for speaker B).

---

1 The groups included 15, 5, 7, and 2 listeners, respectively.

Results

Figs. 13-14 show the scores when the listeners have been treated as one group. Stimuli No. 1 consist of the whole first syllables. A higher number denotes a larger incision.

vå(-ken)

stimulus	1	2	3	4	5	6	7	8	9	10	11
response	76 %	74 %	71 %	77 %	84 %	79 %	71 %	64 %	67 %	64 %	61 %

vå(-ken)

stimulus	1	2	3	4	5	6	7	8	9	10	11
response	71 %	64 %	71 %	68 %	69 %	76 %	67 %	61 %	59 %	66 %	60 %

Fig. 13. Percentages of correct responses to Scanian stimuli (all listeners treated as a single group).

vå(-ken)

stimulus	1	2	3	4	5	6	7	8	9	10
response	16 %	19 %	20 %	27 %	26 %	26 %	44 %	68 %	54 %	65 %

vå(-ken)

stimulus	1	2	3	4	5	6	7	8	9	10
response	21 %	24 %	20 %	19 %	14 %	26 %	26 %	36 %	51 %	50 %

Fig. 14. Percentages of correct responses to West Bothnian stimuli (all listeners treated as a single group).

As for the Scanian speaker (fig. 13) the judgments deviate by +3σ in all cases.

For speaker 8 (fig. 14) the deviation for vå̊ken stimuli No. 1-6 is -3σ, and for No. 7 -2σ. Up to this point it might be said that the listeners are sure, or relatively sure, that the presented word had been vå̀ken instead. Then we get a dramatic shift to positive deviations (+3σ for stimuli No. 8 and 10, while No. 9 only gives +1σ, i.e. guessing).

vàken (for the same speaker) gives -3σ for No. 1-8, then guessing. Positive deviations never occur.

If we split into "dialect" groups the responses are as in figs. 15-18.

Scanian Stimuli

stimulus	1	2	3	4	5	6	7	8	9	10	11
Group 1a	89 %	89 %	83 %	88 %	95 %	92 %	80 %	63 %	69 %	75 %	65 %
Group 1b	55 %	55 %	60 %	70 %	70 %	60 %	50 %	60 %	55 %	40 %	40 %
Group 2	80 %	70 %	68 %	80 %	83 %	70 %	65 %	65 %	68 %	58 %	60 %
Group 3	40 %	47 %	47 %	33 %	53 %	68 %	80 %	73 %	73 %	60 %	80 %

Fig. 15. Percentages of correct responses to Scanian våken.[1]

stimulus	1	2	3	4	5	6	7	8	9	10	11
Group 1a	80 %	75 %	86 %	78 %	83 %	85 %	83 %	72 %	72 %	75 %	63 %
Group 1b	55 %	60 %	50 %	40 %	50 %	45 %	50 %	40 %	40 %	50 %	45 %
Group 2	68 %	55 %	63 %	65 %	63 %	73 %	60 %	58 %	50 %	53 %	58 %
Group 3	47 %	40 %	40 %	53 %	53 %	73 %	40 %	47 %	53 %	67 %	60 %

Fig. 16. Percentages of correct responses for Scanian vàken.[1]

Group 1a (with Scanian listeners) responds in a manner indicating great certainty. The deviation from the theoretical means is in all cases of the order of +3σ. For stimulus No. 11 (våken) this is somewhat surprising considering the fact that the whole fundamental frequency drop during the latter part of the vocoid has been eliminated, and only the initial contoid with its rising fundamental frequency together with a continued fast rise during the very first part of the vocoid remains.[2] For vàken the shortest stimulus also displays a rising f_o but here it is slower. Evidently the rise in itself cannot be the differential cue. Most probably the

1 The groups included 13, 4, 8, and 3 listeners, respectively.

2 For vocoid durations giving different types of responses, see fig. 19.

intensity gives some information. In the acute word the intensity peak is reached, which is not the case in the grave word (cf. fig. 3). This would agree with Hadding-Koch's findings (op.cit., p. 73). She had observed a phase difference between frequency peak and energy maximum for accent 2, while accent 1 on the other hand is characterized by a phase correspondence between these two factors.

Group 1b only in two cases manages to cross the guessing-certainty boundary[1] which is somewhat surprising when we recall the responses to Scanian stimuli in the other tests. As the group is very small, individual listener's responses affect the results to a high degree. Here it is particularly one listener who gives incorrect responses throughout for the Scanian stimuli (and correct for the West Bothnian). In test I the same listener responded quite correctly to both dialects.

Group 2 produces judgments indicating certainty or relative certainty for våken stimuli No. 1-9, then guesses. For vàken guessing appears from stimulus No. 7, where we still have an obvious fundamental rise. This rise is, however, considerably more pronounced for stimuli with lower numbers, where we have certainty.[2]

Group 3 displays great uncertainty all the time which is not very surprising as this group seems to favour an interpretation according to the second syllables.

1 Våken stimuli No. 4 and 5 (+2σ).

2 An "occasional" guessing value also appears for stimulus No. 2.

West Bothnian stimuli

stimulus	1	2	3	4	5	6	7	8	9	10
Group 1a	14 %	15 %	17 %	32 %	22 %	26 %	46 %	60 %	51 %	71 %
Group 1b	25 %	25 %	30 %	35 %	35 %	25 %	50 %	80 %	60 %	70 %
Group 2	10 %	18 %	18 %	23 %	28 %	40 %	50 %	93 %	80 %	83 %
Group 3	33 %	40 %	47 %	33 %	60 %	27 %	53 %	73 %	53 %	53 %

Fig. 17. Percentages of correct responses to West Bothnian váken.[1]

stimulus	1	2	3	4	5	6	7	8	9	10
Group 1a	25 %	23 %	18 %	18 %	9 %	26 %	22 %	29 %	46 %	52 %
Group 1b	20 %	20 %	20 %	20 %	20 %	25 %	40 %	60 %	70 %	70 %
Group 2	20 %	25 %	20 %	8 %	10 %	23 %	23 %	30 %	40 %	38 %
Group 3	7 %	33 %	27 %	47 %	33 %	40 %	40 %	53 %	73 %	47 %

Fig. 18. Percentages of correct responses to West Bothnian vàken.[1]

Scanian listeners (group 1a) give, as was done earlier, acute-responses to grave stimuli, and vice versa. For váken, however, we get guessing for stimuli No. 7 and 9, and acute-responses for stimuli No. 8 and 10. Stimulus No. 7 ends in a small rise, substantially smaller than in No. 6. Evidently a distinct rise is necessary for grave-responses. For vàken we get guessing from stimulus No. 9, where the final frequency drop has been eliminated.

Group 1b reacts very much in the same way as group 1a, and a definite transition to guessing and acute-responses for váken occurs at the same points. Here the listeners are not quite so certain as to the grave character of the word, which is clearly shown in fig. 17. The grave word gives acute-responses as long as the frequency drop is large (stimuli No. 1-6). Then we get guessing for stimuli No. 7 and 8, and grave-responses for No. 9 and 10, where f_o is more level.

1 The groups included 13, 4, 8, and 3 listeners, respectively.

Group 2 needs a substantial rise to give grave-responses (vâken, stimuli No. 1-5). With a more (central and) level f_0 the acute-responses predominate (from stimulus No. 8). The grave word, with an f_0 rise only at the very end of the first syllable, generally achieves acute-responses.

Group 3 reaches certainty only for one stimulus (vàken No. 1) and is in fact sure that this stimulus was vâken instead.

As a complement to figs. 15-18 the vocoid durations for different types of responses are given in fig. 19.

Summary and discussion

The experiments reported above were based on the opinion held by most authors that fundamental frequency differences are the primary cues for the Swedish so-called word tones.

As different dialects may have quite different patterns, two speakers, representing approximately opposite types, were chosen. The listeners, also representing different types, were split into "dialect" groups according to their own fundamental frequency patterns, a more objective means than grouping according to the listeners' personal opinions concerning what dialect they belong to, particularly as some of the dialects represented here evidently constitute transition types between the more extreme patterns used by speakers A and B. This grouping also made it possible to include listeners from dialects with only few representatives.

Test I, where acute and grave words were presented without incisions, clearly shows that we are to a very high degree dependent on our own dialect pattern when we interpret. This is at least the case when we listen to isolated words.

Speaker	Stimulus syllable	I	II				III				IV			
			1a	1b	2	3	1a	1b	2	3	1a	1b	2	3
A	vå(ken)	290 msec	30	–	70	–	–	–	50	–	–	–	–	–
A	vå(ken)	270 msec	55	–	155[1]	–	–	–	135[1]	–	–	–	–	–
B	vå(ken)	230 msec	100	100	120	–	80(40)[2]	80(40)[2]	100	–	60(20)[2]	60(20)[2]	60	–
B	vå(ken)	230 msec	100	140	60[3]	–	80	120	–[3]	–	–	80	–	–

Fig. 19. Vocoid durations: I Full durations, II Minimum durations needed for certainty or relative certainty (negative or positive), III Durations when guessing appears, IV Durations for complete changes in the responses.

1 An occasional guessing appears for stimulus No. 2.

2 Guessing for stimuli No. 7 and 9, and acute-responses for stimuli No. 8 and 10.

3 An occasional guessing appears for stimulus No. 9.

Some listeners seem perceptually to have succeeded in learning more than one pattern. In a few cases the listeners even respond quite correctly to both speakers.

In connected speech the context is bound to give the most relevant information, but also if this should not be the case we probably only need a few words from a speaker to map his pattern in relation to our own. This is one of many areas concerning speaker-listener relations that should be investigated further. From laboratory research we know that acoustic stimuli often give different responses in different surroundings, but we still know very little about what conditions are necessary for us to learn, and to accept, patterns different from the ones we are accustomed to.

Test II, where only the first syllables were listened to, indicates that for groups 1a, 1b, and 2 the information in these syllables is quite enough for a correct identification. For group 3 (with only two listeners in this test, however) the second syllables seem to be of greater importance, but also for the other groups these syllables evidently contain some sort of cue, as the scores particularly for the West Bothnian speaker are considerably lower than in test I.

That the second syllables may be of some importance is clearly shown by test III, especially for the West Bothnian stimuli (speaker B). This is not very surprising, as the fundamental frequency differences between the second syllables of the acute and grave words are rather great, considerably greater than for the Scanian speaker. Judging by Meyer's data there are quite a number of dialects that possess the same great differences, while other dialects do not. (The fundamental frequency differences in these syllables seem to be very small e.g. for the Scanian type, and at least for some dialects of Småland, Halland, Gotland, and Dalarna.) It is, however, evident that the groups 1a, 1b, and 2 to a certain extent succeed

also with the Scanian pattern, in spite of the small f_0 differences, but that they are still more successful with the speaker B stimuli. The small third group guesses for the Scanian stimuli but gives as much as 92 % correct responses to speaker B, i.e. exactly the same as in test I, where the whole words had been presented.

The above observations concerning the distribution of cues between the first and second syllables seem to be corroborated by test IV, where the first syllables of the acute words were combined with the second syllables of the grave words, and vice versa.

It should be remembered, however, that in all our experiments we have been dealing with isolated one or two-syllable stimuli. The second syllables of our test words have weaker stress than the first ones, and, in spite of the results reported here, their importance might decrease in running speech, as the weaker syllables, particularly in accent-1 words, may be used for the sentence intonation (Hadding-Koch, op.cit., p. 66).[1] But it is difficult from the above results to imagine their information load being reduced to zero, at least for dialects with large f_0 differences.

Test V finally, where successive cuts from the ends of the first syllables of våken and vàken had been made, showed, as did the other tests, that the first three groups react in very much the same way. They might consequently be considered to belong to the same "dialect group" with respect to the acute:grave distinction. The scores for the fourth group (group 3 in the tests) do not give very much here, as in this test only the first syllables, or parts of them, were listened to, and in view of the fact that this group throughout had tended to favour an identification according to the second syllables.

---

1 This might, on the other hand help to create greater differences between the second syllables for the two accent types than reported here for the Scanian speaker.

What seems to be quite clear from the scores of the first three groups is that they need a fundamental frequency rise to perceive a grave word,[1] and a fall to hear an acute. The frequency change must sometimes be rather large to be reflected in the listener responses. An exception was stimulus No. 11 for Scanian våken, where the listeners of group 1a (the Scanians) were able to respond correctly already at the frequency peak, i.e. before the typical acute-fall had even started. As cues like duration and vocoid quality hardly can be important here, this probably depends on generally redundant intensity information, perhaps of the kind reported by Hadding-Koch (op.cit., p. 73), i.e. a correspondence between frequency and intensity peaks for accent 1 and a phase difference between these parameters for accent 2.

It is plausible to assume that familiarity with a speaker's voice characteristics is of help when we interpret, as is probably the case for other phenomena than the word accents as well, but Fintoft-Mårtony (op.cit.) suggest that the subjects probably do not relate fundamental frequency variations to the average frequency used by the speaker, but merely identify the accents by the f_o variation itself. In 1965, however, Fintoft assumes that the starting point of the fundamental frequency in the vocoid in relation to the average fundamental frequency used by the speaker is a cue for the identification. This may very well be true, particularly if the starting point is very different from the average fundamental frequency so that one of the accent possibilities may be effectively excluded. There is at least some evidence to this effect in this test. But a rise or a fall of the fundamental frequency seems to be the essential thing.

If the fundamental frequency in the first parts of the vocoids should be

---

1 Cf. Fintoft-Mårtony (1964) who were working with listeners that needed a fall for this accent, instead.

of primary importance, this would be contrary to the findings of e.g. Heinz et al. (1967), where the last part of a stimulus (with continually changing frequency) was proved to play a dominant role. This has, furthermore, been shown to be valid also for more complex stimuli (Johansson, 1967, 1969a, and 1969b, concerning final formant transitions).

That the relation between various parts of an utterance may sometimes be of importance was shown by Jassem (1963), who only by altering the fundamental frequency of a preceding unstressed syllable brought about change in the responses.

Malmberg (1955) felt inclined to regard the place of a frequency peak within the stressed vocoid and the direction of the melody curve following from this difference as the distinctive feature. With the peak in the first part of the vocoid (for the Scanian dialect) accent 1 was heard, while a peak in the latter part achieved a change to grave responses. For other types of Swedish the pattern may be quite the opposite.

Test V above showed that the position of the peak in itself cannot always be the distinctive cue, but rather the fundamental frequency change, upwards or downwards.

In spite of the often continually changing fundamental frequencies, and in spite of what has been said about level frequencies not giving the impression of word tone, I would prefer, like Malmberg (1962), to speak of low and high levels from a perceptual point of view - low for the acute with falling frequency, and high for the grave with rising, at least for the dialect regions covered by groups 1a, 1b, and 2. One reason for this is the way in which group 2 reacts in this test. Acoustically, both the acute and the grave words for this group are characterized by mainly falling fundamental frequencies in the first vocoids, the difference being that for the grave words the fundamental starts at a higher frequency, and hangs

on there longer, than for the corresponding acute words. In spite of the listeners' own acoustic patterns, they need, as did the listeners of groups 1a and 1b, a rather pronounced rise to interpret a stimulus as grave, i.e. it seems as if their own acoustically high (level) grave type could be regarded as a variant, an "allotone", of the rising type. An investigation of how listeners from the other dialects would have responded to a speaker from the group 2 region would probably have enlightened us further on this point.

There is evidently no doubt that the fundamental frequency of the stressed syllable is a dominant cue for the acute:grave distinction, at least for some dialects, but the experiments above have also shown that the second syllables may be important. (This must, however, be investigated further.) It also seems probable that the intensity may provide sometimes essential redundant information. Other factors, as duration, vocoid quality, can hardly be primary cues, nor can the syllable boundary.

However, one interesting thing is that most parameters discussed in the literature in connexion with word tones are parameters that are able to contribute to the perception of loudness. On the functional level the whole thing may therefore very well in the end turn out to be a question of prominence or stress, at least for some people. This would help to explain why whispered "word tones" are possible, as reported by some authors (e.g. Segerbäck, 1966). The more similar frequency patterns of the two syllables within the grave words for some dialects, and the relations of these patterns to each other, are factors that may also very well fit such a way of reasoning.

References

Efremova, I.B. - Fintoft, K. - Ormestad, H. (1963), Intelligibility of tonic accents, Phonetica 10, p. 203-212

Elert, C.-C. (1970), Ljud och ord i svenskan

Fant, G. (1958), Modern instruments and methods for acoustic studies of speech, Proceed. 8th congr. ling., p. 282-358

Fintoft, K. (1965), Some remarks on word accents, Phonetica 13, p. 201-226

Fintoft, K. - Mártony, J. (1964), Word accents in East Norwegian, STL-QPSR 3, p. 8-15

Gjerdman, O. (1954), Acc. 1 och acc. 2, akut och gravis, Nysvenska studier 32, p. 125-154

Hadding-Koch, K. (1961), Acoustico-phonetic studies in the intonation of southern Swedish

- (1962), Notes on the Swedish word tones, Proceed. 4th congr. phon., p. 630-638

Haugen, E. - Joos, M. (1952), Tone and intonation in East Norwegian, Acta Philologica Scandinavica 22, p. 41-64

Heinz, J. - Lindblom, B. - Lindqvist, J. (1967), Patterns of residual masking for sounds with speech-like characteristics, STL-QPSR 2-3, p. 66-71

Jassem, W. (1963), An experiment in word accent perception with synthetic speech, STL-QPSR 1, p. 10-12

Jensen, M.K. (1961), Tonemicity, Acta Univ. Bergensis, Ser. Hum. 1

Johansson, K. (1969 a), Försök avseende vokaltransitions riktning och dess betydelse för "plats"-distinktionen bland tonande klusiler, Working Papers 1, Phonetics Laboratory, Lund University

- (1967), Några perceptoriska försök med svenska dentaler och supradentaler, Thesis (unpublished work)

- (1969 b), Transitions as place cues for voiced stop consonants: direction or extent?, Studia Linguistica XIII:II

Malmberg, B. (1959), Bemerkungen zum schwedischen Wortakzent, Zeitschrift für Phonetik und allgemeine Sprachwissenschaft 12:1-4, p. 193-207

Malmberg, B. (1961), "Haut ~ Bas" ou "Montant ~ Descendant"?, Language and Society; Essays presented to Arthur M. Jensen to his seventieth birthday, p. 99-107

- (1962), Levels of abstraction in phonetic and phonemic analysis, Phonetica 8, p. 220-243

- (1965), Observations on the Swedish word accent, mimeographed report from Haskins Laboratories

- (2nd ed. 1967), Structural linguistics and human communication

- (1953), Sydsvensk ordaccent, Lunds Universitets Årsskrift

Meyer, E.A. (1927), Die Intonation im Schwedischen I

- (1954), Die Intonation im Schwedischen II

Segerbäck, B. (1966), La réalisation d'une opposition de tonèmes dans des disyllabes chuchotés

STL-QPSR = Speech Transmission Laboratory. Royal Institute of Technology, Stockholm. Quarterly Progress and Status Report

Öhman, S. (1967), Word and sentence intonation: a quantitative model, STL-QPSR 2-3, p. 20-54.