PHONETICS LABORATORY
LUND UNIVERSITY

WORKING PAPERS
3 · 1970

PHONETICS LABORATORY
LUND UNIVERSITY

WORKING PAPERS

3 · 1970

DIPHTHONGIZATION IN THE MALMÖ DIALECT

Gösta Bruce

1. Definition of the Malmö dialect. The dialect under examination in the present study constitutes a part of the Scanian dialect ("skånska"), the kind of Swedish used in the very south of Sweden. Malmö is the principal town of the province of Scania ("Skåne") with about 250 000 inhabitants, i.e. one fourth of the population of Scania. Scanian itself exhibits in certain important respects phonetic characteristics quite different from those of Standard Swedish. This is in part due to the fact that the province is former Danish territory and only in the 17th century - after the conquest of Scania by Sweden - did Swedish become the official language. The particular dialect used in this part of Sweden can thus be said to be Swedish on Danish substrate. This means that, with some over-simplification, the "code" is identical or quasi-identical with that of Standard Swedish, while some of the habits of pronunciation are Danish, or to be more precise, East-Danish. Since the conquest Scania has become increasingly Swedish. According to Witting (1959) the following general remarks can be made about the Scanian dialect from a phonetic point of view: "Its three chief characteristics of pronunciation are, briefly: (1) the exceptionless use of the uvular (or velar) r instead of Central Swedish alveolar r; (2) a tendency towards a pervading diphthongization; (3) an almost complete inversion of the intonational pattern of tonal accents, "acute" (accent I) versus "grave" (accent II) as found in Central Swedish". The present study will be confined to the examination of point 2, i.e. the diphthongization. (A brief discussion of the definition of the concept of diphthong will be given below. So far I use the term 'diphthong' operationally in the sense of diphthongized vowel.)

2. The status of the diphthongs. To avoid misunderstanding I will make a distinction between two types of diphthong in Scanian. The Malmö diphthongs (as well as those of other towns in southern and western Scania) are very recent developments. They are opposed to the diphthongs found in northern Scania both in pronunciation - although they have certain common features - and in historical development. These north Scanian diphthongs developed in the 15th century (Wessén). It is the latter type of diphthong that is usually referred to in the literature as typically Scanian. In the present Study I will concentrate upon the Malmö type.

When examining the diphthongs one notices a social distribution in their use. Generally speaking, the diphthongization of the vowels is regarded as socially substandard and is avoided in the "educated" dialect, where there is a tendency to approach the Standard Swedish norm by using more or less "pure" vowels. It is, however, outside the scope of this study to penetrate this "vertical" dimension too deeply. The diphthongs in the Malmö dialect are regarded merely as surface manifestations of single underlying phonemes. The diphthong has no direct distinctive function (See for further discussion p. 5: 'short vs. long vowels'). Each of the diphthongs is represented in orthography by one single grapheme. The "naive" speaker of the dialect - and indeed often the educated speaker - is not usually conscious of his own diphthongized vowels.

3. The vowel system in the Malmö dialect - i.e. the relations between the vowels contained in the system (disregarding the physical manifestation) is very similar to that of Standard Swedish. There are 9 vowel phonemes

in opposition to each other, all 9 of which appear long and 7 short (cf. Stand. Sw. 9 long vs. 9 short). Vowel length rather than vowel quality is considered to be distinctive in Stand. Sw., although there may be simultaneous differences in vowel quality. (See Hadding-Koch – Abrahamson, Elert, Lindau for discussion.) For the moment we accept this view for the Malmö dialect too. The following vowels are found (the same symbols are used for long and short vowels, ignoring at this level any quality differences between a long-short vowel pair):

long = /i:/ /e:/ /ɛ:/ /y:/ /ø:/ /ʉ:/ /u:/ /o:/ /ɑ:/

short = /i/ /e/ /y/ /ø/ /u/ /o/ /ɑ/

As shown above there are gaps for short /ɛ/ and /ʉ/. The choice of the symbol /e/ instead of /ɛ/ to cover the gap is arbitrary. It does, however, reflect the phonetic reality. The coalescence of short /e/ and /ɛ/ is a common phenomenon in Swedish dialects. The short /ʉ/-gap, however, is less common. Modern Swedish /ʉ/ has developed from an earlier back /u/-vowel. This fronting has also taken place in the Malmö dialect, giving long /ʉ:/, but the short counterpart has never been fronted.

4. The vowel target. So far I have used diphthong in the sense of diphthongized vowel. As mentioned above there are assumed to be no underlying or phonemic diphthongs in the Malmö dialect. The diphthongized vowels are the physical manifestations of single phonemes.

Generally speaking the vowel phoneme – in any language – can be said to represent the vowel target, the ideal phonetic value at which the speaker aims. Each vowel, – syllable nucleus – has at least one target. The target can be manifested differently in the time dimension; (i) as one steady state throughout the vowel (monophthong, i.e. a more or less static vowel), (ii) in the initial part only, followed by an off-glide from the target steady state or (iii) in the final part only preceded by an on-glide to the target steady state. Type (ii) and (iii) (Lehiste-Peterson [1961] "glide" but often in every day usage referred to as "diphthongs" or "diphthongized vowels") are dynamic as opposed to the first type. Further, there are (iv), two-target vowels or diphthongs, ideally manifested as two steady states linked by an intermediate glide. I shall not go further into the problem of defining the concept of diphthong at this superficial level, as the question does not seem to be crucial for the present study.

In the diphthongal realizations of the long vowel phonemes in the Malmö dialect, it is the second part of the syllable nucleus that constitutes the target; i.e. the target is reached with a time delay (type iii). The diphthongization lies so to say in the initial part of the vowel; there is usually a relatively long on-glide, sometimes even beginning with a steady state position (See fig. 1!). Whether to classify the diphthongized vowels in the Malmö dialect as diphthongs or glides - with reference to the above definition - is an open question. Some of them may be interpreted as having two targets, others seem to have just one. If you claim that a steady state is a manifestation of a target, then some of the vowels in the Malmö dialect could no doubt be called diphthongs (type iv). It is probably futile to make this decision on phonetic grounds only. ["Glide" according to the definitions (ii) and (iii) given above is from now on included in the term "diphthong".]

5. Effect of speaking rate. I believe that it is a question of speaking rate, whether the targets are manifested as steady-states or not. One must take into account that it is only in optimal situations, that a diphthong is realized substantially as two steady states with an intermediate glide connecting them. This optimum is seldom reached in spontaneous speech. Thomas Gay (Haskins) has recently investigated the effect of speaking rate on diphthongs in American English. In JASA (Dec. 1968) he writes: "Results

indicate that onset target position and second formant rate of change are fixed features of the diphthong formant movement, while offset target positions are variable across changes in duration." If this statement proves to be true not only for American English it would be possible to describe a diphthong acoustically in terms of onset steady state position and especially second) formant rate of change with an indication, whether it is a plus- or minus-transition. At a fast speaking rate the offset target is never reached. In articulatory terms: the tongue is moving at the same rate from one vowel position towards another independently of speaking rate. The accuracy of hitting the target positions is just a question of the speaking rate.

6. Thriphthongs. It has been suggested that there are triphthongs in the Malmö dialect. Malmberg (Svensk fonetik 1968) makes the claim that triphthongs are common in the southern Swedish dialect and that it is possible to discern even more vowel hues within one syllable nucleus. This statement is however not supported by any experimental data available to me for the Malmö dialect. Of course this may depend upon the definition of "triphthong". Malmberg has used a perceptual definition: three clearly descernable vowel qualities in one syllable nucleus. Thus when the vowel (diphthong) is for some reason being lengthened, e.g. in emphatic pronunciation, there is a greater chance to descern more vowel qualities. But in my opinion it is just complicating the description of the dialect in question to introduce the concept of triphthong on perceptual grounds only.

7. Short versus long vowels. So far I have discussed the diphthongization of the long vowel phonemes. With respect to their phonetic manifestation the 7 short vowels in the Malmö dialect can be transcribed as [i] [e] [Y] [ø] [u] [o] [æ][1] (See fig. 2). The short vowels - at least the more close

1 The short vowels in Malmö speech are not entirely identical in quality with the Stand. Sw. counterparts, which I have tried to express by the use of other phonetic symbols for some of them. For example [Y] is less palatalized and [æ] is more front.

ones, are also sometimes diphthongized. In the short vowels the diphthongization lies in the final part of the vowel (type ii) as opposed to that of the long vowels. The target is reached at once, after which there is, or may be, a gliding off the target towards a more open and central vowel psition.

Preliminary measurements of the quantity of "short" and "long" vowels suggest that in the Malmö dialect there is little or sometimes no difference in length between them (cf. for Stand. Sw. - Elert 64). I would say that the difference - in the substance - lies rather in the diphthongization: the "long" vowels are manifested as diphthongs - with an initial onglide to the target; the "short" vowels are manifested as monophthongs or may be diphthongized with an offglide from the target, which is in the initial part of the vowel. In this sense the diphthong (diphthongized long vowel) may be attributed a distinctive function versus the monophthong (or differently diphthongized short vowel - See page 2 -). A minimal pair bot /bu:t/ (= penance) ---> [be͡ut] - bott/but/ (= lived ---> [bu͡ət] may exemplify the statement. The difference in the Malmö dialect is thus primarily one of diphthongization.

Another striking fact - obvious from the spectrographic analysis - is that the short vowels compared to the long vowels are usually more "extreme", their position on the vowel diagram is further away from the neutral vowel position than those of their long counterparts. (I am referring to the second part of the long diphthongized vowels.) Tense - lax (See Jakobson-Fant-Halle 1952) reversed in the Malmö dialect? Further investiagtion of this point may show whether this is due only to the speaking-rate causing undershooting of the target. It may shed some light o the definition of the tense - lax concepts.

8. Diphthongization. As has been mentioned above all of the long vowel phonemes are more or less diphthongized - at least in the "broad" dialect.

(In the "educated" speech the diphthongization is reduced, but even if people try to reduce the diphthongization in accordance with the Stand. Sw. norm, it is striking that the vowels /u:/ and /o:/ very often remain diphthongs in their speech.) If we look more closely at the actual diphthongization of each vowel, we will easily find some regularities. Spectrograms of the diphthongized vowels are shown in figure 1. The speaker is a high school student (aged 19) and can be said to be representative of the common Malmö dialect, i.e. neither the extremely broad nor the "educated" dialect. From the spectrographic data a vowel diagram – F1–F2 plot – (figure 2) has been made. Each diphthong is marked with an arrow indicating the direction in which the diphthong glides. The short vowels are used merely as references and are marked with a small cross, the crossing-point of F1–F2. The findings are also supported by studies of spectrograms – kindly put to my disposal by a colleague – of about 50 Malmö subjects with a similar background.

One regularity apparent from the spectrograms is that the first component of each diphthong is more compact than the second component. (The terms used here are defined in Jakobson et al. 1952.) A further observation is that the more diffuse the diphthong ends the more diffuse is the beginning. A third observation is that the first part of the diphthong is always acute, independently of the features of the second part. When comparing the two groups of acute vowels – the first group having a plain target, and the second having a flat target – the latter group also starts flat compared with the former group. If we reinterprete these findings in articulatory terms, they would be as follows: the first component of each long diphthongized vowel (diphthong) in the Malmö dialect is (i) more open than the second component – the opening degree being dependent of that of the second part, (ii) non-back and (iii) usually non-round with the exception of the front rounded vowels.

9. Grave vowels. You will make an interesting discovery if you examine the vowel diagram (fig. 2) a little more closely. The diphthong that is the realization of the phoneme /u:/ has an F2 that starts higher than that of any other vowel (F3, however, is comparatively low). This observation is evident from the overwhelming majority of spectrograms I have examined. The originally grave vowel has a more acute starting-point than any of the acute vowels, In accordance with Gay's theory of the effect of speaking rate on diphthongs mentioned above, the vowel /u:/, produced at a fast speaking rate would not reach the offset target, consequently loosing some of its grave (or velar) quality. Thus the vowel will become even more acute. How much of its velar quality can /u:/ loose without the perceptual result being affected? The same question is applicable to /o:/, which also has an acute starting-point. It is plausible that a rather insignificant glide from the onset target towards the offset target is sufficient for a perceptual identification of /u:/ or /o:/. It would be of great interest to make perceptual tests in order to see how much of the final component of the diphthong be eliminated without the perceptual identification being lost. A transcription - to a great extent supported by the spectrographic data available - of the diphthongized /u:/ and /o:/ would give [e͡u] and [ɛ͡o]. The interpretation of the diphthongization of the vowel /ɑ:/ as [a͡eɑ] or even extremely [a͡eo] is in my opinion more adequate than the traditional [a͡u]. (Malmberg 1968.)

10. Acute vowels. The vowel /i:/ is usually diphthongized [e͡i]. A frication of the final element of the diffuse vowels /i:/, /y:/, /ʉ:/, and /u:/ is sometimes found, but is not as emphazised in Malmö speech as in Stand. Sw. (See Lindblom). The diphthongized /i:/ found in fig. 1 displays

a pattern that is more representative of the "educated" Malmö dialect than of the common dialect. The vowel is somewhat "Viby-coloured" (for definition see Malmberg). The vowel /e;/ can be transcribed phonetically [ɛ͡e]. The final component has a very high F2 and is sometimes interpreted as [i]. The vowel /ɛ:/ is less diphthongized than the other acute vowels and may be transcribed [æ͡ɛ]. The vowel in question exhibits a vowel quality that is peculiar for Malmö (and Southern Scanian) speech: a flattening that might be due to pharyngeal constriction. It would be interesting to use X-ray film to investigate this question. Some of the same quality as in /ɛ:/ is present in /ø:/, diphthongized as [œ͡ø]. This latter vowel ends more compact than Stand. Sw. /ø:/. The transcription of the diphthongized /y:/ should be [ø͡y]. The final part of the diphthong is not necessarily [y], but may also be [i] or [j]. The initial part, however, must be [ø] (flattened) to give a correct identification of the vowel. The vowel /ʉ:/, finally, has about the same starting-point as /y:/, i.e. the transcription is [ø͡ʉ]. To sum up the transcriptions:

/i:/ ⟶ [e͡i]	/y:/ ⟶ [ø͡y]	/u:/ ⟶ [e͡u]
/e:/ ⟶ [ɛ͡e]	/ʉ:/ ⟶ [ø͡ʉ]	/o:/ ⟶ [ɛ͡o]
/ɛ:/ ⟶ [æ͡ɛ]	/ø:/ ⟶ [œ͡ø]	/ɑ:/ ⟶ [æ͡ɑ]

11. Neutral vowel theory. An explanation of the historical development of the diphthongs in the Malmö dialect is not easy to give and will be only speculative. I will point to a few facts that may be interesting in this connection. Regarding the vowel diagram (fig. 2) it is apparent that each diphthong (with perhaps one exception: the diphthongized /ø:/) starts at a fairly neutral position and ends at a more extreme position. This is a further generalization of what has been said above about diphthongization. I believe, however, that it is reasonable to say that all Malmö diphthongs theoretically start at a neutral position (schwa).

This statement implies that the more extreme the vowel the greater will the difference be between the first and the second component of the diphthong. This seems to be supported by the spectrographic data to a certain extent. In fact, then an adaptation of the first to the second component has taken place, reducing the risk of perceptual confusion; i.e. the close vowels have a relatively closer beginning than the open vowels and the front rounded vowels are rounded not only in the final part but already in the beginning of the diphthong to contrast more effectively with their unrounded equivalents. But the back vowels - redundantly rounded because they have no back unrounded equivalents - do not run this immeidate risk of confusion, so the first part is not necessarily rounded. The existence of an epenthetic vowel before the target in the long vowels, resulting in a diphthong, might be regarded as a coarticulation phenomenon. The vowel gesture is quite different in the Malmö dialect compared with Stand. Sw., where we find a well developed coarticulation within the syllable. Prevocalic consonants in the Malmö dialect are not appreciably influenced by following long vowels; in other words a non-synchronizing of the consonantvowel gesture might have contributed to the development of the Malmö diphthongs. It, should, however, be noted that even in a V(C)-syllable, i.e. when a vowel is syllable initial, this vowel will become a diphthong. Therefore if the diphthong is regarded as a product of coarticulation, this fact can only be explained by analogy. The CV(C)-syllable is more common, so the V(C)-syllable has adopted the same pattern.

12. Diphthongization rules. Finally I will try to formalize in phonological rules within the general framework of generative phonology the observations about the diphthongization in the Malmö dialect presented

above (See for explanation of formal devices 'Sound Pattern of English' ch. 8). The generalizations to be captured in the rules should express at least the following facts (to repeat what has been said above).

First: the diphthong consists of a glide segment - it may also be called a semi-vowel or a non-syllabic vowel p that is inserted before the original (long) vowel segment -- the vowel phoneme, i.e. the syllabic vowel:

$$\text{(i)} \qquad \emptyset \longrightarrow \begin{bmatrix} -\text{ syl} \\ -\text{ con} \end{bmatrix} / \;\_\_\_\; \begin{bmatrix} +\text{ syl} \\ -\text{ con} \\ +\text{ long} \end{bmatrix}$$

The feature [+ long] here implies the feature [+ stress]. In an unstressed position the long vowels are not diphthongized. Rule (i) may also be formulated in an alternative way, namely:

$$\text{(ii)} \qquad \begin{bmatrix} +\text{ syl} \\ -\text{ con} \\ +\text{ long} \end{bmatrix} \longrightarrow \begin{bmatrix} -\text{ syl} \\ -\text{ con} \end{bmatrix} \quad \begin{bmatrix} +\text{ syl} \\ -\text{ con} \end{bmatrix}$$

The content in rule (i) and (ii) is almost the same. Rule (ii) expresses that the long vowel splits up into two segments, where the feature [+ long] appears as the segment $\begin{bmatrix} -\text{ syl} \\ -\text{ con} \end{bmatrix}$, which may be a more elegant way of expressing the facts. To recapitulate: a diphthong - in the Malmö dialect - is defined as a semi-vowel + a vowel.

The next step is to find a proper framework for the description of the different vowel segments that are contained in the diphthongs. Languages that use four opening degrees for vowels (Malmö speech included) are not easily described either within the Jakobsonian (1952) or the Chomsky-Halle (1968) feature system. Wang (Language, 4 1968) uses only two features - $[\pm\text{ high}]$ and $[\pm\text{ mid}]$ to account for four tongue heights. Accepting Wang's feature framework, the description of the "front" group of vowels will be as follows:

		i	e	ɛ	æ
(iii)	high	+	+	–	–
	mid	–	+	+	–

To account for three places of production (front, central and back vowels) Wang uses the features palatal and velar with the redundancy rules:

(iv) [+ palatal] ⟶ [– velar] and
[+ velar] ⟶ [– palatal]

If we choose for example the vowels [y] [ʉ] [u] that represent three different places of production, a matrix would have the following appearance:

		[y]	[ʉ]	[u]
(v)	pal	+	–	–
	vel	–	–	+

I believe, however, that it would be correct to have an alternative version of Wang's redundancy rules, yielding instead:

(vi) [– pal] ⟶ [+ vel] and
[– vel] ⟶ [+ pal]

and thus the matrix:

		[y]	[ʉ]	[u]
(vii)	pal	+	+	–
	vel	–	+	+

The treatment $\begin{bmatrix} + \text{pal} \\ + \text{vel} \end{bmatrix}$ to account for the Swedish [ʉ]-vowel has been proposed by B. Lindblom (QPSR 1, 1969)[2]. I will adopt this analysis here. In this connection it should be noted that the inserted vowel segment – the semi-vowel – always has the feature [+ pal], or even [– vel] to restrict it even more. This is the <u>second</u> generalization to be made about

2 The Swedish [ʉ] is phonetically a "front" vowel, but might at this level of representation be defined as $\begin{bmatrix} + \text{pal} \\ + \text{vel} \end{bmatrix}$ (for further discussion see Lindblom op. cit).

the diphthongization process.

I will use the feature [± labial] to distinguish between the two groups of vowels that are produced with protruded or spread lips respectively.

A binary matrix of the vowel segments that are part of the Malmö diphthongs – using Wang's features – is given below:

(viii)

	[i]	[e]	[ɛ]	[æ]	[y]	[ø]	[œ]	[ʉ]	[u]	[o]	[ɑ]
pal	+	+	+	+	+	+	+	+	–	–	–
vel	–	–	–	–	–	–	–	+	+	+	+
high	+	+	–	–	+	+	–	+	+	+	–
mid	–	+	+	–	–	+	+	–	–	+	+
lab	–	–	–	–	+	+	+	+	+	+	+

As there are only three different tongue heights for the velar vowels in the dialect, I have defined the vowel [ɑ] as $\begin{bmatrix} -\text{high} \\ +\text{mid} \end{bmatrix}$, which would have been the definition of [ɔ] if four tongue heights were represented for the back vowels.

Now in combining the vowel segments in the matrix to form diphthongs, the following <u>third</u> generalization can be formulated: An assimilation process with respect to tongue heights is taking place. The inserted half-vowel adopts an opening degree that is one degree higher than that of the original "mother" vowel, It is possible to formalize this process using either binary or n-ary features, The binary model is formulated in (ix):

$$\text{(ix)} \quad \begin{bmatrix} +\text{syl} \\ -\text{con} \\ +\text{long} \end{bmatrix} \longrightarrow \begin{bmatrix} -\text{syl} \\ -\text{con} \\ +\text{pal} \\ -\beta\ \text{high} \\ \alpha\ \text{mid} \end{bmatrix} \begin{bmatrix} +\text{syl} \\ -\text{con} \\ \alpha\ \text{high} \\ \beta\ \text{mid} \end{bmatrix}$$

For example if /i:/ is the original vowel with the features $\begin{bmatrix} +\text{high} \\ -\text{mid} \end{bmatrix}$ the glide segment goes to $\begin{bmatrix} +\text{high} \\ +\text{mid} \end{bmatrix}$, i.e. [e], giving the diphthong [e͡i] (we assume so far that the glide becomes [– lab]). Using graded features - n-ary features[3] - instead of binary features to account for the tongue heights the rule would appear alternatively as (x):

$$(x) \quad \begin{bmatrix} + \text{ syl} \\ - \text{ con} \\ + \text{ long} \end{bmatrix} \longrightarrow \begin{bmatrix} - \text{ syl} \\ - \text{ con} \\ + \text{ pal} \\ (n+1) \text{ high} \end{bmatrix} \quad \begin{bmatrix} + \text{ syl} \\ - \text{ con} \\ n \text{ high} \end{bmatrix}$$

Replacing the features [± high] and [± mid] by graded height, such that 1 refers to the highest vowel, 2 to the next highest and so on, we have

		[i]	[e]	[ɛ]	[æ]	[y]	[ø]	[œ]	[ʉ]	[u]	[o]	[ɑ]
(xi)	high	1	2	3	4	1	2	3	1	1	2	3

So if the mother vowel has tongue height 1, the instrusive element has height 2; for example [e] - height 2 - will be preceded by [ɛ] - height 3 -, [ɛ] by [æ] etc.

The n-ary treatment to account for the tongue heights of vowels seems to be a simpler and more direct way than the corresponding binary treatment. The "one step up"-principle that is present in the Malmö diphthongization process does illustrate this. A complement to the ~~genuine~~ binarity - binary treatment at the systematic phonemic level (for terminology see Harms 1968), then rewriting of the binary units into integers in the low level rules at the systematic phonetic level. It is not clear, where in the hierarchy of phonology to insert the diphthongization discussed above. Is it a superficial rule close to the final phonetic output, ir is it a deeper-lying process? A correct evaluation of the methods of treating tongue heights of vowels for the Malmö dialect presupposes an answer to this question.

Four: regarding the labialization of the vowels there is an assimilation process going on too. The front mother vowels influence the inserted semi-

3. In the rules (x) and (xiii) [n] may assume one of the values 1, 2, 3.

vowels with respect to the labialization, whereas their back equivalents do not. Therefore we will have to divide the vowels into two groups: a front (palatal) and a back (velar) group. Taking into account the labialization process (disregarding the vowel height for the moment) the rule appears as (xii):

$$\text{(xii)}\qquad \begin{bmatrix}-\text{ syl}\\ -\text{ con}\\ +\text{ pal}\end{bmatrix} \longrightarrow \left\{\begin{matrix}[\alpha\text{ lab}] \;/\; \text{—} \begin{bmatrix}+\text{ syl}\\ -\text{ con}\\ +\text{ pal}\\ \alpha\text{ lab}\end{bmatrix} \\ \\ [-\text{ lab}] \;/\; \text{—} \begin{bmatrix}+\text{ syl}\\ -\text{ con}\\ -\text{ pal}\end{bmatrix}\end{matrix}\right\}$$

This labialization process may, however, be expressed otherwise by saying that before an unmarked vowel $\begin{bmatrix}\alpha\text{ vel}\\ \alpha\text{ lab}\end{bmatrix}$, e.g. [e] or [o], the glide becomes unmarked, i.e. [– lab], because the glide has the feature [– vel]. In contrast a marked vowel, for example [ø], influences the glide to become marked too, i.e. [+ lab] in this case.

In rule (xiii), which is an attempt to compress in one rule all general facts concerning the diphthongs expressed so far, we find that:

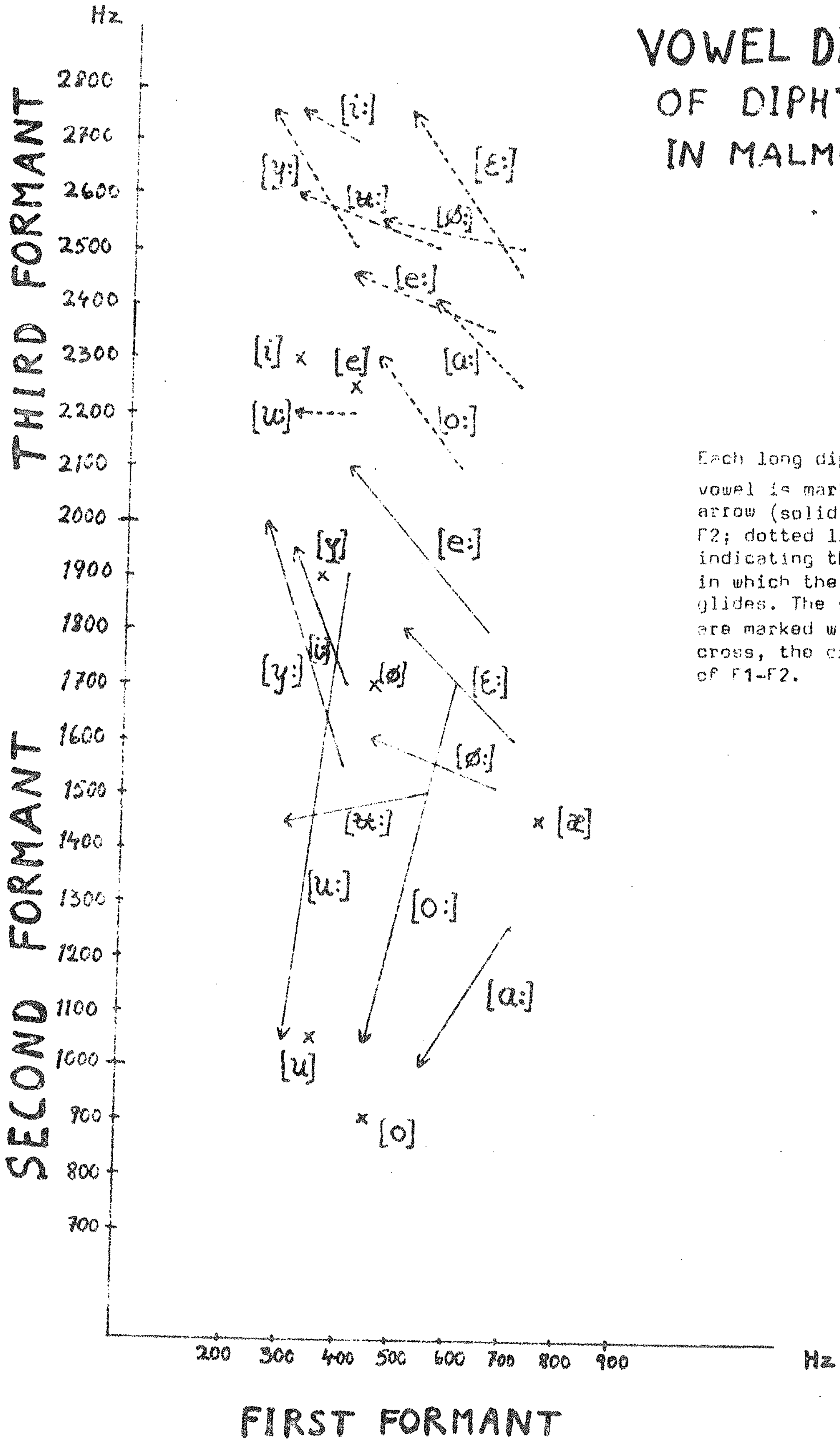
$$\text{(xiii)}\ldots\emptyset \longrightarrow\ldots \begin{bmatrix}-\text{ syl}\\ -\text{ con}\\ +\text{ pal}\\ (n+1)\text{high}\end{bmatrix} \;/\; \left\{\begin{matrix}\begin{bmatrix}\text{—}\\ \alpha\text{ lab}\end{bmatrix} \;/\; \text{—} \begin{bmatrix}+\text{ pal}\\ \alpha\text{ lab}\end{bmatrix} \\ \begin{bmatrix}\text{—}\\ -\text{ lab}\end{bmatrix} \;/\; \text{—}\; [-\text{ pal}]\end{matrix}\right\} \;/\; \begin{bmatrix}\text{—}\\ +\text{ syl}\\ -\text{ con}\\ +\text{ long}\\ n\text{ high}\end{bmatrix}$$

This rule generates the optimal diphthongs [e͡i], [ɛ͡e], [æ͡ɛ], [ø͡y], [œ͡ø], [ø͡ɵ], [e͡u], [ɛ͡o], [æ͡ɑ], which seems to be correct. Output constraints (not yet formulated) will then operate on this rule to give the final phonetic product. Further, it will be required that the diphthognization process be related to other rules in the phonological component (e.g. stress and quantity rules), which may affect some of the features in the rule (xiii).

It should be emphazised that the present study is only preliminary. It will be used as a working hypothesis for further research. Therefore comments will be much appreciated.

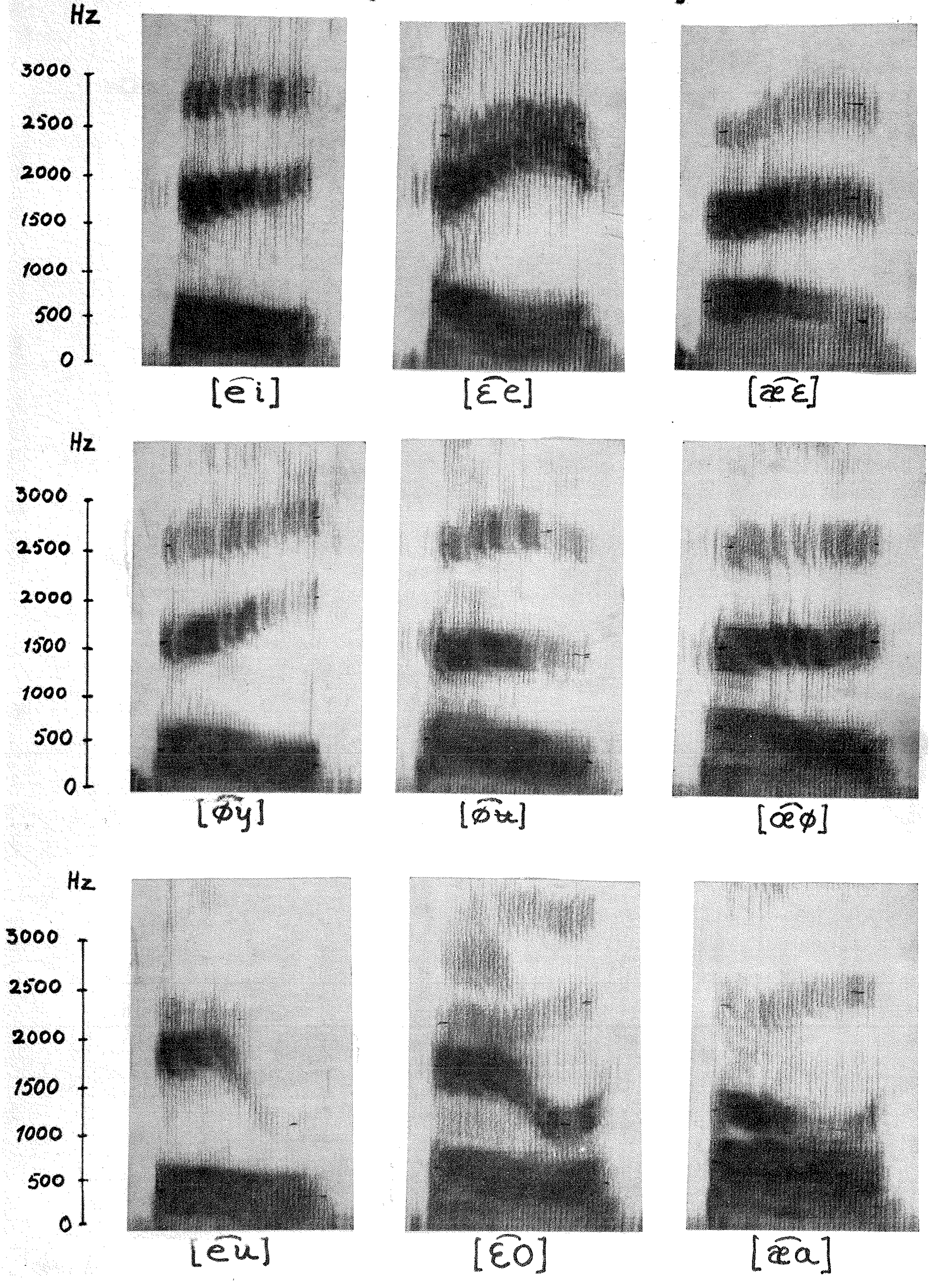
Fig. 2

VOWEL DIAGRAM OF DIPHTHONGS IN MALMÖ SPEECH



Each long diphthongized vowel is marked with an arrow (solid line = F1-F2; dotted line = F1-F3) indicating the direction in which the diphthong glides. The short vowels are marked with a small cross, the crossing-point of F1-F2.

Fig. 1 Spectrograms of diphthongized vowels.



References

Chomsky, N., & Halle, M., 1968, The sound pattern of English. New York.

Elert, C-C., 1964, Phonologic studies of quantity in Swedish. Stockholm, Uppsala.

Gay, Th., 1968, Effect of speaking rate on diphthong formant movements. JASA, 14:6, 1570-1573.

Hadding-Koch, K., & Abrahamson, A., 1966, Duration versus spectrum in Swedish vowels: some perceptual experiments. Studia Linguistica, 18:2, 94-107. Lund.

Harms, R.T., 1968, Introduction to phonological theory. Englewood Cliffs, N.J.

Hovdhaugen, E., 1970, Transformasjonell generativ grammatikk. Oslo.

Ingers, I., 1957, Språket i Lund (Skrifter uts. av Nämnden f. sv. språkvård, 17). Stockholm.

Jakobson, R., Fant, G., & Halle, M., 1952, Preliminaries to speech analysis. Massachusetts Institute of Technology. Acoustic Laboratory, Technical Report, 13.

Lehiste, I., & Peterson, G., 1961, Transitions, glides, and diphthongs. JASA 33:3, 268-277.

Lindau, M., 1970, Prosodic problems in a generative phonology of Swedish. WP 2. Phonetics Laboratory, Lund University.

Lindblom,B., & Sundberg, J., 1969, A quantative model of vowel production and distinctive features of Swedish vowels. STL-QPSR 1.

Malmberg, B., 1968, Svensk fonetik. Lund.

Wang, W., S-Y., 1968, Vowel features, paired variables, and the English vowel shift. Language 44:4, 695-708.

Wessén, E., 1958, Våra folkmål. Stockholm.

Witting, C., 1959, Physical and functional aspects of speech sounds. (Uppsala Universitets årsskrift 1959:7) Uppsala.

WORD TONES AND LARYNX MUSCLES

Eva Gårding

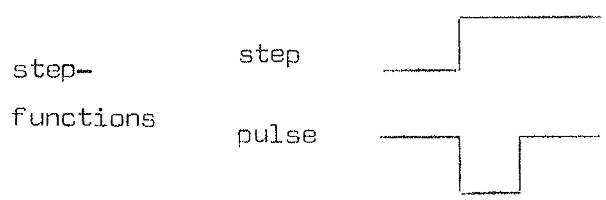
0. Outline

This is a slightly revised version at a seminar talk given in Lund, May 1970. In Section 1 I shall make a brief summary of Öhman's model for intonation (Öhman, 1967 a) and in Section 2 I shall discuss the results of an EMG study that was carried out by Öhman, Leandersson, and Persson to test Öhman's theory (Öhman et al. 1967 b). Section 3 is a preliminary report on an EMG investigation parallel to Öhman's (See also Gårding, Fujimura, Hirose, 1970). The implications of our own data are discussed in Section 4. They indicate that Öhman's model cannot be entirely correct. In Section 5, a revised version is suggested and in 6, Meyer's wordtone data for Swedish dialects (Meyer, 1937 and 1954) are interpreted in terms of the revised model. A correlation of our EMG data to the standard stress notation used for the word tones concludes the report (7).

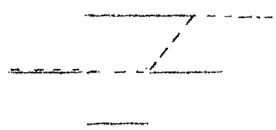
1. Summary of Öhman's model

Öhman tried to incorporate the Scandinavian word tones in a general quantitative model for intonation (Öhman, 1967 a). The first version of this model only concerns the stressed syllable of an utterance. He postulates that the observed fundamental frequency curve for such a syllable can be decomposed into a positive sentence intonation step and a negative word intonation pulse.

To give an example:



The stepfunctions represent the on and offset of the nerve signals. Each stepfunction passes through a filter which simulates the inertia of the anatomic structures. The filter has a smoothing effect on the stepfunction. When the two inputs above have been smoothed and added to each other, the output will be like the dotted line below:



By timing the pulse in various ways in relation to the step, Öhman obtains the configurations that he needs to approximate the word tones in different dialects. (Meyer, 1937 and 1954. The dialectal variation is shown in Figure 1, Öhman's figure II-B-3, op. cit. p. 23.)

Dialectal variation can then be explained as a difference in timing between the pulse and the step. The almost reversed patterns in Stockholm and Skåne (Numbers 1-5 and 90-93 in Figure 1) are interpreted in the following way: The time order of the step and the pulse is reversed (op. cit. p. 26). The acute and grave accents correspond in Danish to stød (a kind of glottal stop) and no stød. Smith (Smith, 1944) made acoustic and physiological measurements of this contrast. In an EMG investigation of the expiratory muscles he found that words with stød are characterized by a brief and intense innervation pattern compared to a more evenly distributed innervation in the words without stød. This innervation is followed by a sudden relaxation that (for reasons not well understood) brings about a momentary disturbance or inhibition in the vibration of the vocal cords. The effect of this inhibition is a segment with a glottal stop or in most cases a segment with creaky voice.

The fact that a glottal stop in Danish corresponds to the acute accent in Swedish gave Öhman the idea of interpreting all Scandinavian word tones as "glottal stops" in the sense that they all have some inhibitory effect on

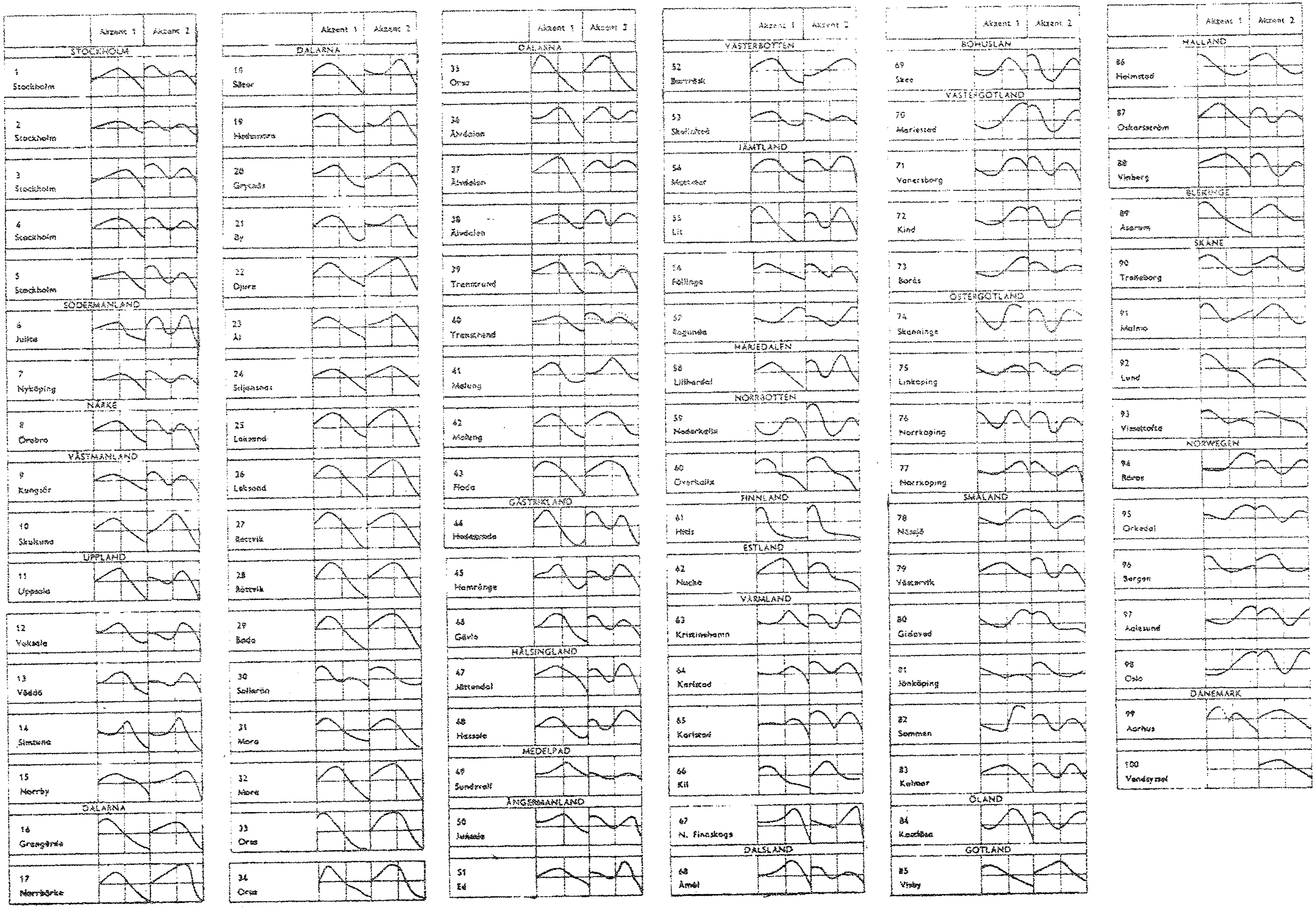


Figure 1.

Fig. II-B-3. Schematic acute and grave accent patterns of a hundred Scandinavian dialects according to E. A. Meyer: Die Intonation im Schwedischen, part II.

sentence intonation. This inhibitory effect is strongest in Danish: for the stød the vocal cords are adducted to such an extent that the vibratory pattern becomes irregular. In other Scandinavian dialects - according to the theory - the effect is much weaker: the vibrations are inhibited but they do not lose their periodicity.

This glottal stop theory motivates Öhman's choice of a negative pulse for the word accent.

2. Öhman's EMG study

Öhman then looked for some physiological mechanism that was likely to have a negative, that is pitch inhibiting effect on sentence intonation. In an EMG investigation Öhman and his co-workers studied the behavior of two of the intrinsic laryngeal muscles during the pronunciation of words with contrastive word tones and words with glottal stops. (Öhman et al., 1967 b.) The vocalis muscle was chosen as a representative of the adductory muscles. It is natural that this muscle should be active for glottal stops. That this is indeed the case has been shown in earlier EMG investigations, for instance by Faaborg-Andersen. (Faaborg-Andersen, 1957.) The cricothyroid was selected because this muscle is known as the main tensor muscle and therefore the muscle responsible for pitch control. When the cricothyroid contracts the cricoid arch is raised towards the anterior part of the thyroid cartilage. In this process the arytenoid cartilages which are mounted on top of the cricoid lamina are tilted backwards. The tilt lengthens the distance between the points of attachment of the vocal cords. (Sonesson, 1968, p. 51.) The vocal cords will therefore be stretched which means that during phonation the pitch will go up. The effect of the cricothyroid activity on pitch has been extensively studied. (Arnold 1961, Faaborg-Andersen 1957, Ohala et al. 1968 and 1969.) Expected electromyographic correlates to Öhman's theory would be a period of inhibition in the cricothyroid muscle combined with increased activity in the vocalis at the time when the negative word pulses are supposed to occur.

The results of the EMG experiment, which was performed with needle electrodes on a male subject speaking the Stockholm dialect, were inconclusive. The main finding according to Öhman was a brief interval of inactivity in the cricothyroid muscle which roughly coincided with the negative pulse. There was no corresponding increase in vocalis activity, at least not in the examples published by the authors. Figure 2 shows examples of the vocalis and cricothyroid activity implied by the word tones. The inhibition phases related to the word tones are marked by arrows.

To facilitate the interpretation of the electromyograms I have added the fundamental frequency curves that can be expected in this kind of dialect. The cricothyroid activity correlates very well with the fundamental frequency curve in general. With a low pitch as in the last part of the frame ja sa there is a concomitant low activity in the cricothyroid. In my view this correlation can also explain the arrow marked interval in the grave accented word in the figure and is not necessarily an indication of a negative pulse. Increased activity in the vocalis would have made the negative pulse interpretation more probable but the vocalis record does not change in connection with the pulse.

The electromyograms of the acute accented word do not give much support for the negative pulse either. It is true that the arrow points to an interval with reduced activity in the cricothyroid signal at a point in time tha fits the postulated pulse but it is not accompanied by increased activity i the vocalis muscle, rather the contrary. In my opinion the reduced activit in both the cricothyroid and vocalis muscles could be an effect of a phonation pause in connection with the syntactic boundary between the frame de v and the test word rather than a manifestation of a negative pulse. The gap

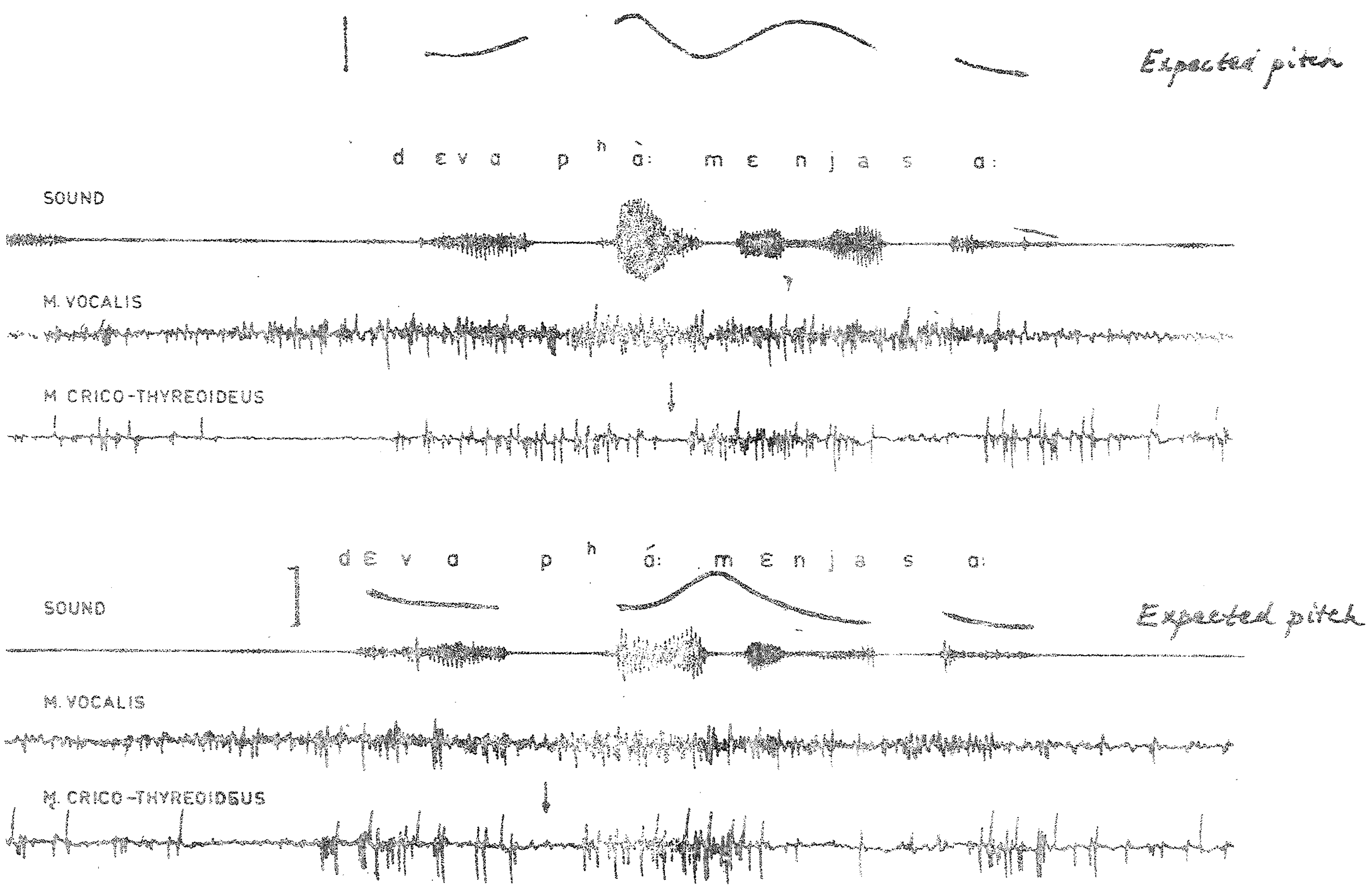


Figure 2. Expected fundamental frequency curves have been added.

Fig. II-C-1. Examples of vocalis and crico-thyroid motor unit activity during the production of acute and grave accented words. The crico-thyroid inhibition phase related in the oscillogram for the occlusion of the p of the acute word is longer than normal and certainly quite a bit longer than the occlusion for the corresponding p above in the grave accented word.

This experiment is the only attempt that Öhman has made so far to give a physiological basis to his theory.

In the following I shall discuss some results of an EMG investigation parallel to Öhman's. On the basis of our EMG data I shall propose an analysis of word tones in terms of positive pulses only. This analysis is well supported by the EMG data. It also fits the word tone data collected by Meyer.

My analysis may also be used to sketch the development of word tones in Scandinavian dialects from a common origin. I shall come back to this in another context.

3. EMG study (Gårding, Fujimura, Hirose, 1970)

Subjects

The two subjects are E (female) representing the Skåne dialect and L (male) speaking a dialect in which the word tones are manifested much the same as in the Stockholm dialect. His dialect may perhaps be descrited as Standard Middle Swedish.

Test material

Our test material consists of some 20 sentences with the two word tones in varying phonetic contest. The test words were put in neutral frames and the sentences were read with and without emphasis in a rising-falling sentence intonation. A typical test sentence is de va pamen ja se, i.e. it was pamen I said. The nonsense sequence pamen was given the acute and the grave accents respectively. A few items had glottal stops and a couple of sentences were whispered. Each test sentence was read fifteen times serially. So far only part of the material has been processed.

Muscles

Since the experiment was a parallel to Öhman's, it was natural to choose the same muscles: the vocalis and the cricothyroid. In addition the sternohyoid was selected. This muscle has been shown to be active for pitchlowering in both English and Japanese (Hirano et al. 1967 and Ohala et al. 1968) and it was expected to have a similar function in Swedish.

EMG equipment

Figure 3 shows the EMG equipment. The electrodes are bipolar and consist of thin wire threaded through the cannula of a hypodermic needle. The end of the wire is hooked. When the desired location has been reached the needle is withdrawn leaving the electrode hooked to the muscle. (The experimental procedures are described by Hirose et al., 1970.)

The mucosa of the larynx is given a light anesthesia before the electrodes are inserted through the neck. The insertion is almost painless. Once the electrodes have been fixed in the right position the subject can talk in a natural voice without feeling their presence. After the session the wire is removed by a light tug.

The electromyographic signal was amplified by a DC preamplifier designed for EMG. It was recorded together with the microphone signal on magnetic tape. The recorded signals were fed to a PDP-9 computer through an AD (analogue to data) converter. In this process the EMG signal is sampled every 100 microsec and digitized into 6 bit levels. At first absolute values were taken of the samples and these were then smoothed over a range of 10 msec. (For a better description, see Simada et al. 1970.) The smoothed signals for 8-12

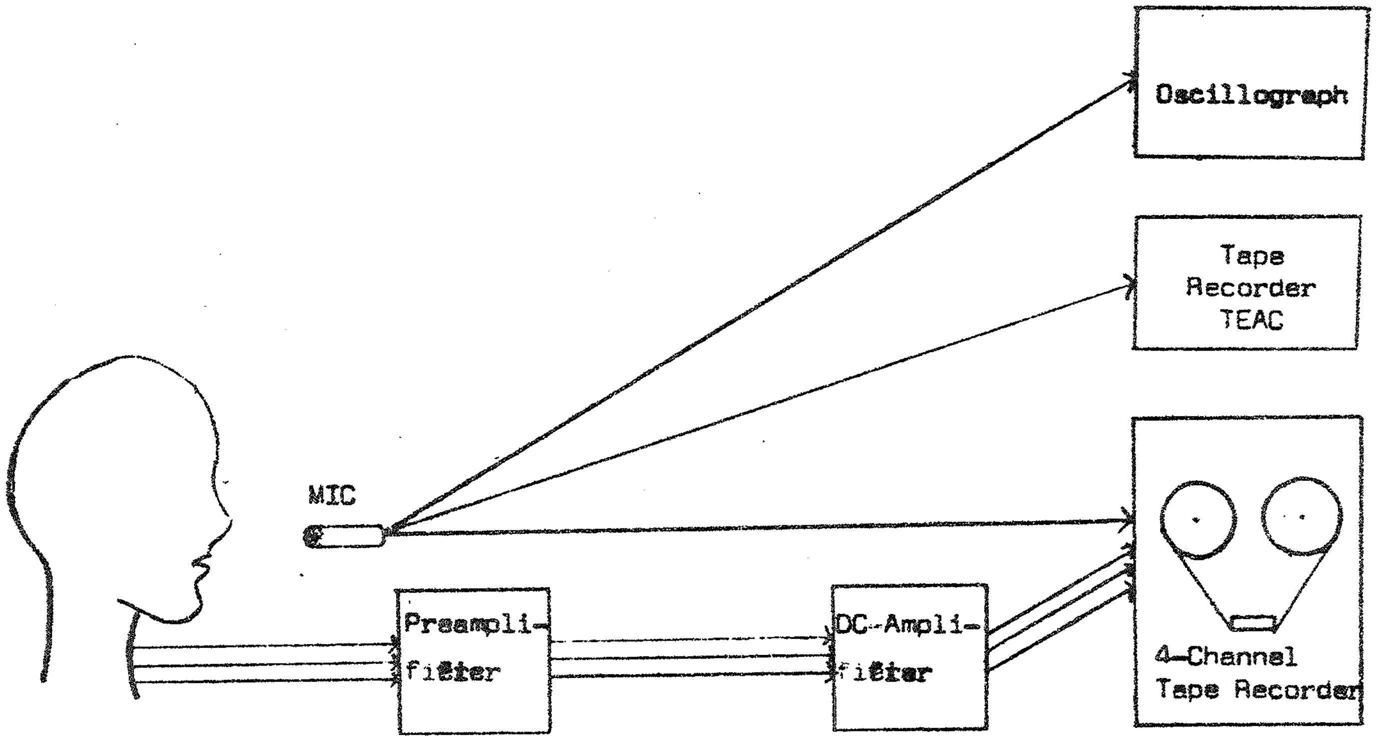


Figure 3. EMG equipment

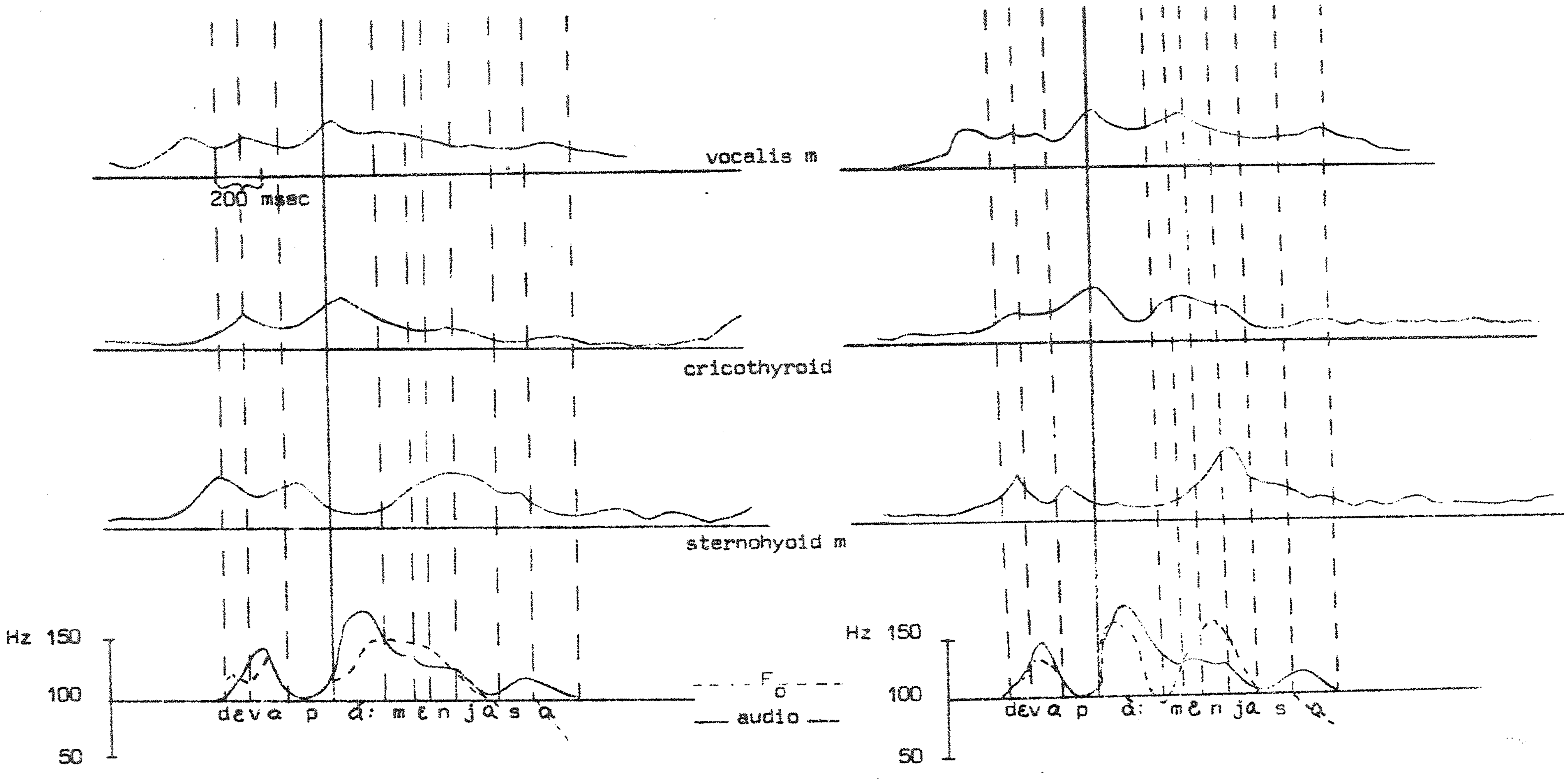


Figure 4. Speaker L. The EMG and Audio signals represent 12 rectified averaged and smoothed utterances. The F_o curve (dotted line superimposed on the audiosignal) is from one representative utterance.

ing the larynx back to its rest position after having been lowered for the low-pitched end of the frame? Or is it an effect of the inspiration? Buchthal (1959) thought that an increase of activity in this muscle during inspiration might serve to steady the vocal cords and prevent them from flapping during the influx of air.

The cricothyroid and the fundamental frequency curves are very similar. Every peak in F_0 has a corresponding peak in the cricothyroid curve. This confirms what is well known from a number of earlier investigations: the cricothyroid muscle is the main pitch controller. The articulatory gestures have little or no effect on this muscle.

The sternohyoid has peaks when the jaw opens for the most open and the most stressed vowels.

Figure 5 brings out the contrast between the acute and the grave accents, pá:mɛn, pàmɛn in 5 a and mámma, màmma in 5 b. The vowel segments of the acute words are marked by thick lines along the base line and the vertical lines show the borders of these segments. The corresponding lines for the vowel segments in the grave accented words have been left unmarked to avoid cluttering up the figure. They would in each case precede the existing lines by a few millimeters.

The most important difference between the word tones in the vocalis and cricothyroid behavior is two peaks for the grave accent compared to one for the acute, that is in principal the same difference as we find in the resulting fundamental frequency curves. The first peaks connected with the grave accent occur a little earlier than the acute peaks. The sternohyoid activity obviously is of no importance for the word tone contrast. The audio signal is similar for the two accents and refutes the traditional view of the grave accented words as having different energy distributions from the acute ones.

At the top of the figure Öhman's pulses have been inserted at the place on

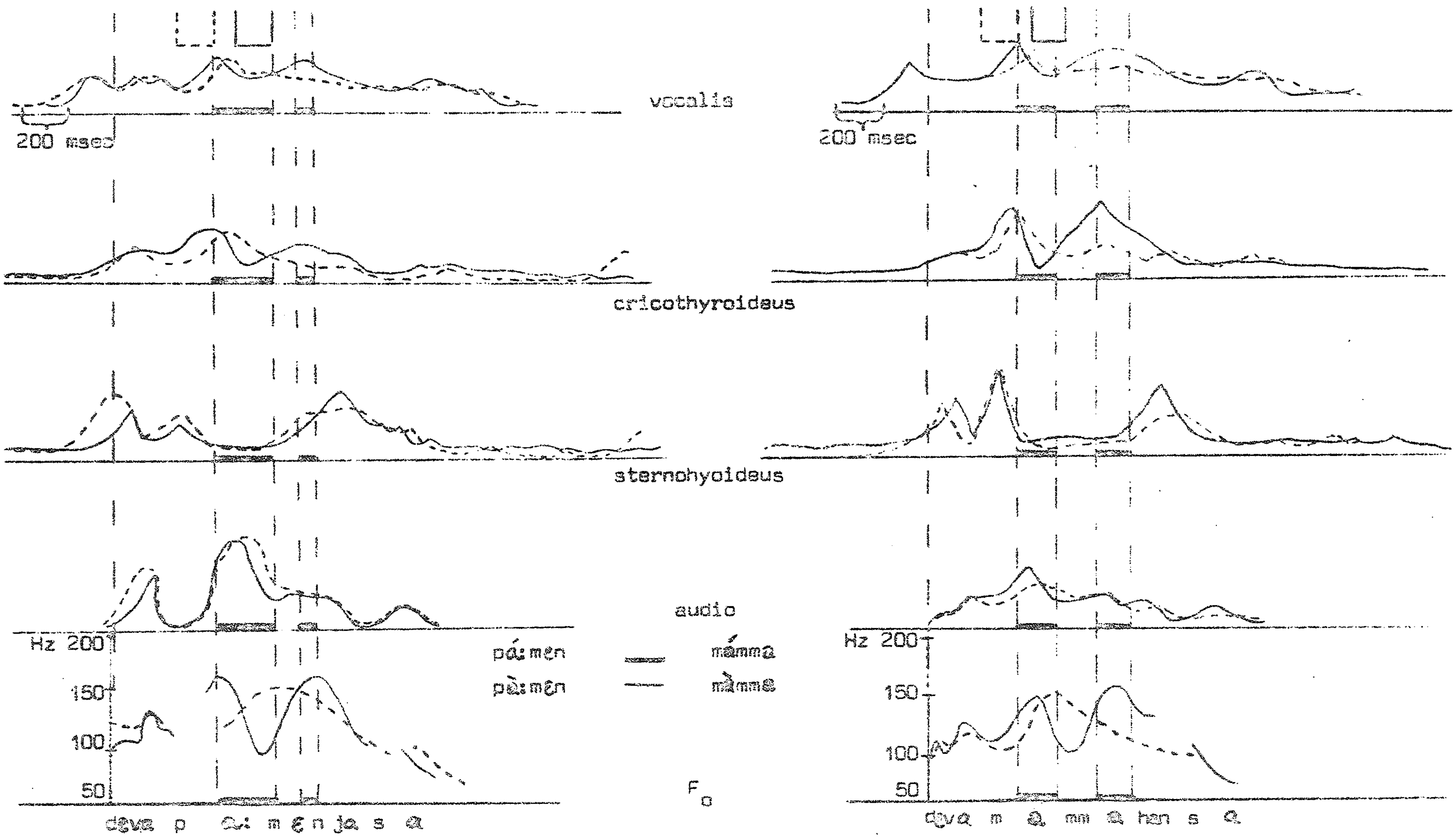


Figure 5. Speaker L. Rectified averaged and smoothed signals from two pairs of test sentences with contrastive wordtone. Tracings from utterances with the acute accent have dotted lines, the grave accent utterances have straight lines.
5 a: pá:mən, pà:mən
5 b: mámma, màmma.
Vowel segments of the acute words are marked by thick lines along the base line. Vertical lines show borders of these segments.
Öhmans pulses are inserted at the top of the figure at the place in time postulated for this dialect. Dotted pulse is the pulse of the acute accent, straight lined pulse is used for the grave accent pulse.

the time scale where they have been postulated for this dialect: early for the acute (dotted line) and late in the first syllable for the grave (straight line).

To support his theory the vocalis -- as an adductor -- would have to be active during the pulse whereas the cricothyroid - as a tensor - would have to be inhibited. The most convincing argument against Öhman's pulse theory in these records is that the vocalis and cricothyroid muscles work together. Both muscles contract for the pitch peaks and they reduce their activity for the pitch fall in the accented words as well as for the lowpitched end of the frame.

There is - in these data - no resemblance between the muscle activity controlling word tones and the activity needed for the production of a stød or a glottal stop.

The utterances behind Figure 6 contain what can be regarded as extreme cases of glottal stops. On the whole these utterances need a maximal use of laryngeal control because they have, apart from a glottal stop between the two syllables, the grave accent and emphatic stress.

The utterances represented in these tracings are to the left [jàʔa] and to the right [nɛ̀:ʔɛ] pronounced by Speaker E in the Skåne dialect. The most striking feature in these tracings is perhaps that when the vocalis is active for the glottal stop, that is the closure of the glottis, the cricothyroid is completely relaxed.

What is the mechanism behind the opposed activity in the vocalis and cricothyroid muscles in this kind of glottal stop? Are these muscles opposed for adduction in general and working together for pitch control i.e. tensing? The data obtained so far speak in favour of such an interpretation. Notice again the pronounced peak in the vocalis muscle with a corresponding valley in the cricothyroid before the beginning of phonation. This event is similar

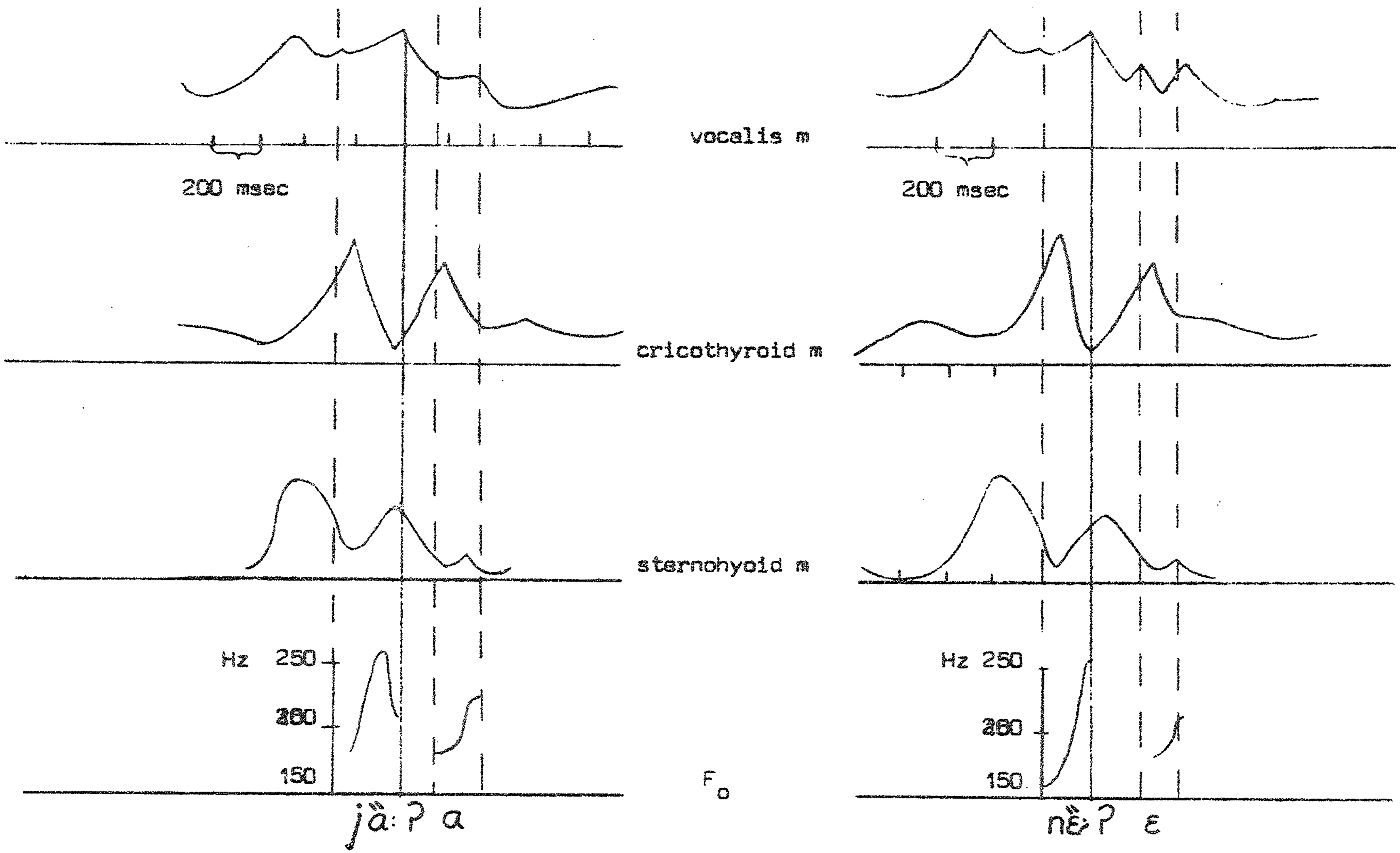


Figure 6. Speaker E. The signals represent 8 rectified, averaged and smoothed utterances of the test sentences [jȁ:ʔa] to the left and [nɛ̏ʔɛ] to the right. They are emphatic renderings of the words ja and nä (yes and no). The double grave accent symbols stand for the emphatic grave accent.

to the muscle activity for the glottal stop and may very well be a related adductory gesture.

In the corresponding test sentences for Speaker L, the glottal stop in most of the utterances is pronounced in a different way: no burst can be heard at the beginning of the second vowel. In the EMG tracings this performance of the glottal stop is parallelled by a less pronounced dip in the cricothyroid curve. It seems that the degree of relaxation in the cricothyroid is dependent on the degree of tension in the vocalis muscle.

Figure 7 presents some more data from Speaker E. You notice the typical Skåne accent patterns in the fundamental frequency curve: The acute accent is manifested by a rise and fall in the stressed syllable and the grave accent by a rise. As can be expected, this difference from Speaker L is accompanied by a difference in the EMG tracings.

The vocalis and cricothyroid muscles do not follow each other as closely as they did in Speaker L. They do have similar main peaks but the vocalis, apart from a deep valley for the occlusion of p, has a number of smaller dips that can be related to the other consonants. (Notice the opposed activity in the vocalis and cricothyroid muscles before the beginning of the audio signal.) The main peaks of the cricothyroid muscle coincide with peaks in the fundamental frequency curve.

The sternohyoid muscle has peaks when the jaw opens for the syllables. The highest peak is connected with the finishing syllable sa. This may be an indication of the extra activity needed to bring down the larynx in the appropriate position for the relaxed vocal cords needed for the low tone of the end of the frame.

Figure 8 shows the contrast between the acute and grave accents in Speaker E. Dotted lines represent the acute accent and the grave accent is marked by straight lines. Öhman's pulse is added as before. The difference between the

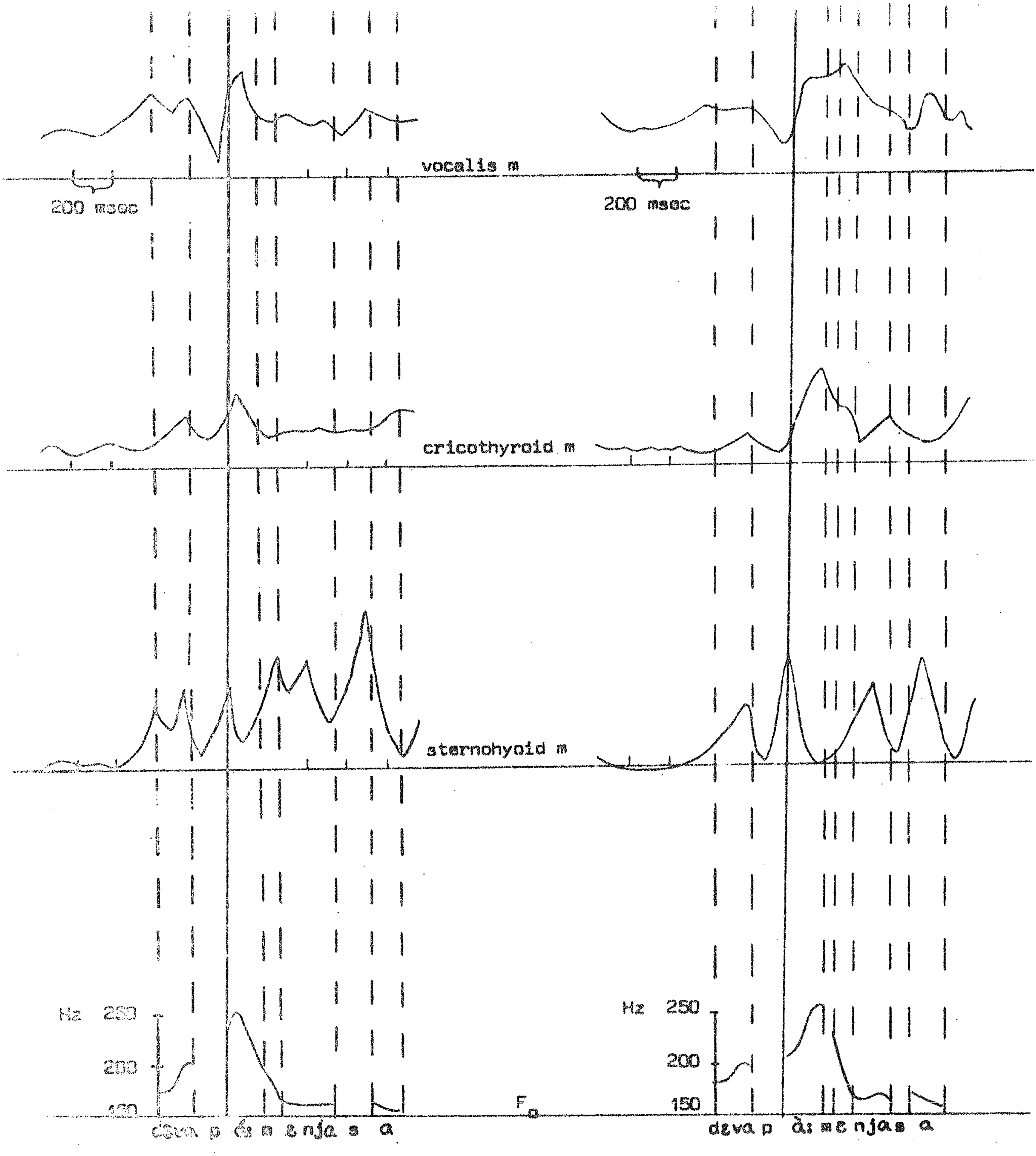


Figure 7. Speaker E. Arrangement of signals as in Figure 5.

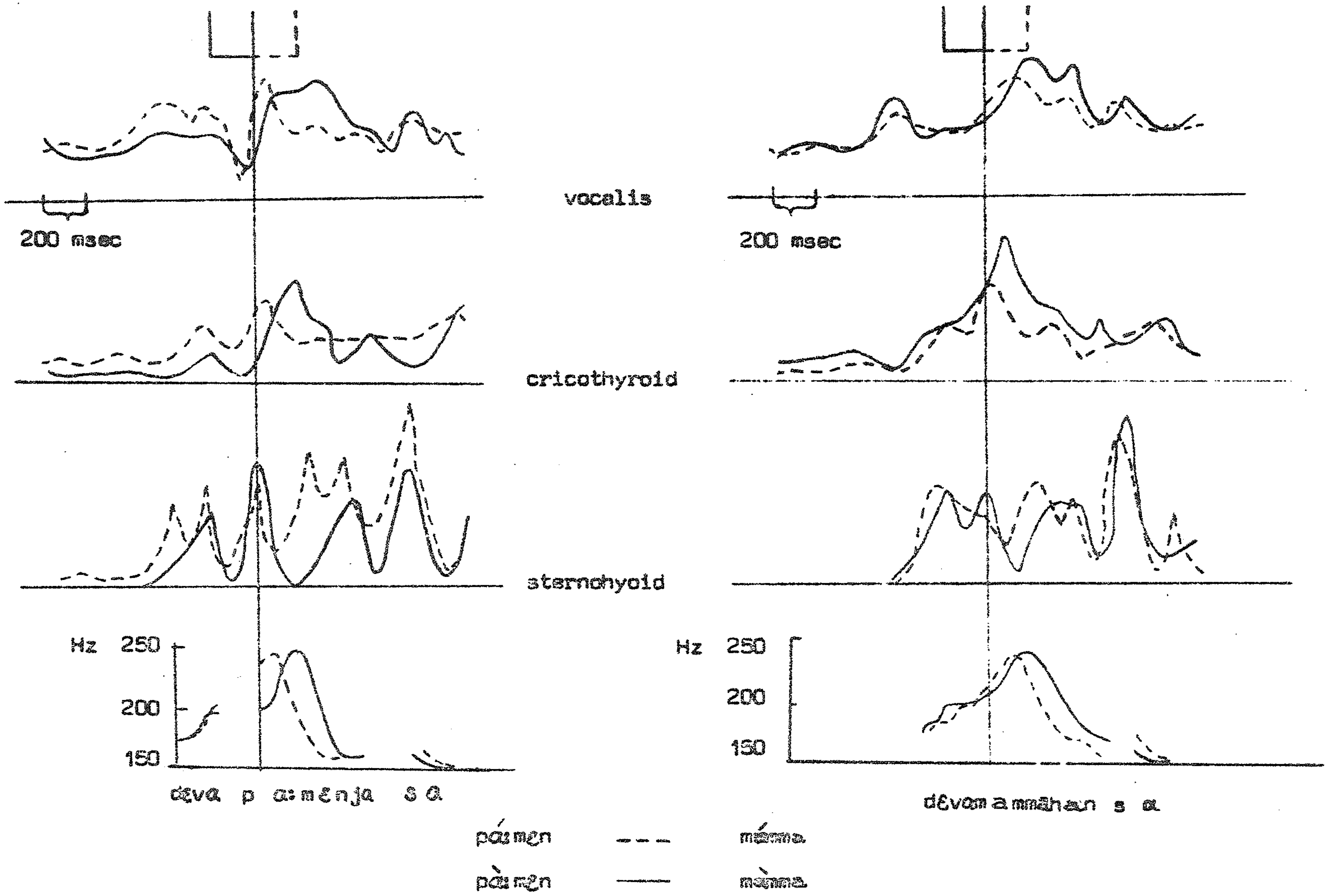


Figure 8. Speaker E. Word tone contrasts in _pamen_ to the left and _mamma_ to the right. Arrangement of signals as in Figure 5. The pulses have been added at the place in time postulated for this dialect, i.e. early for the grave (straight lines) and late for the acute (dotted lines).

word tones as they appear in these records is an early peak in the fundamental frequency, the vocalis and the cricothyroid curves for the acute accent as opposed to a later and plateau-formed obtrusion for the grave accent. According to the theory we should expect to find increased activity in the vocalis concomitant with reduced activity in the cricothyroid. As can be seen in the figure the muscles do not behave in this way. In three cases out of four the vocalis and the cricothyroid muscles cooperate. Only in mámma (acute) is there for a short period during the pulse opposed activity in the two muscles.

To sum up, our data do not provide any physiological basis for Öhman's negative pulse.

4. Discussion

Öhman was of course well aware that his EMG results were inconclusive, but he suggests that if the EMG investigation included other laryngeal muscles as well the pitch inhibiting muscle would perhaps be discovered.

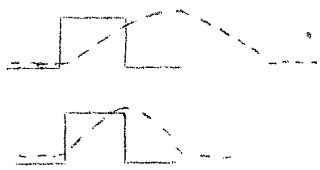
Is there any physiological justification for such a view?

It appears from the discussion on p.29 (op.cit.) that Öhman expects muscular activity for the word tone to be "ballistic" in nature just as Smith – using Stetson's terminology – regarded the stød as a ballistic stroke (Smith, 1944 b, p. 3). A typical innervation pattern behind such a movement would be an activation of an agonist muscle combined with a complete inhibition of its antagonist (reciprocal inhibition). In the EMG records from the test words pronounced with glottal stops (Figure 6) the opposed activity in the vocalis and the cricothyroid may perhaps be due to what is called reciprocal inhibition. But there is no similar effect in the data that can be associated with the word tones. Since there is no activity in the vocalis during the word tone pulses, Öhman's interpretation of the acute and grave accents as some kind of glottal stops is not very convincing.

5. Word tones and positive pulses

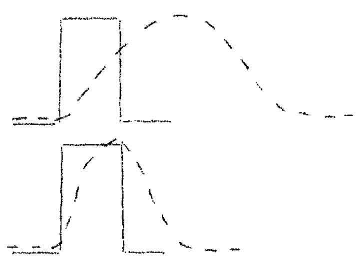
In the light of our data it seems natural to simulate word tones by one or two possitive pulses. To fit an actual fundamental frequency curve, these pulses must have appropriate amplitudes, onset times and durations and they must be exposed to a suitable smoothing procedure. I shall give a few examples of what can be achieved by such a process:

same pulse,
different degrees
of smoothing



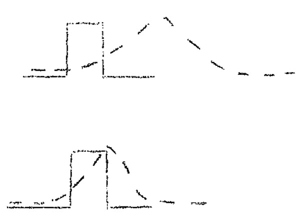
slow smoothing
fast "

pulse with high
amplitude and
different degrees
of smoothing



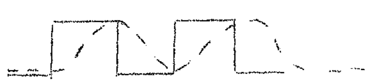
slow smoothing
fast "

pulse with short
duration and
different degrees
of smoothing



slow smoothing
fast "

two pulses
fast smoothing

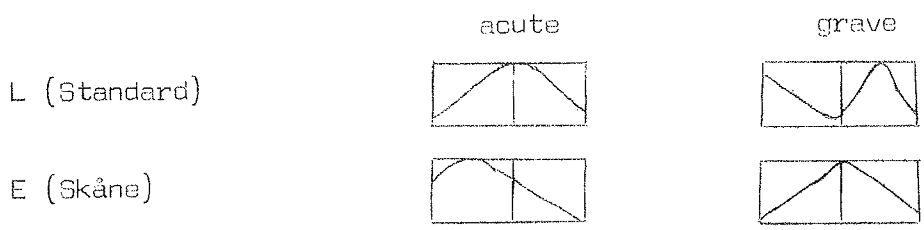


It seems clear that with one or two positive pulses we can simulate any pitch pattern implied by the word tones.[1] Actually in my examples above I may have more parameters than I need but we may safely say that a pitch curve with one peak corresponds to a step function with one pulse (or perhaps two very close pulses) and a pitch curve with two peaks corresponds to a step function with two pulses.

From now on we shall consider curves that are similar in shape to the obtrusions caused by the pulses in my examples as the output of positive step functions that I shall call positive pulses.

We have found earlier that the EMG curve derived from the cricothyroid muscle has about the same shape as F_0 with a time delay typical of an acoustic signal compared to a physiological one. The similarity between the F_0 and the cricothyroid curves makes it possible to interpret obtrusions in the fundamental frequency curve as the visible representatives of positive pulses.

Let us now go back to the word tones in the dialects of my speakers and interpret them as positive pulses.



The pitch curves have been stylized in the manner of Meyer with one square for each of the two syllables.

In terms of positive pulses the difference between the acute and grave accents in Speaker L is one pulse for the acute and two for the grave, whereas

---

1 According to Lars Gårding it is an established mathematical fact that any smooth curve can be simulated by a step curve provided that the step curve is subjected to a smoothing process with fixed parameters.

E has an early pulse in the acute as compared to a late pulse in the grave accent.

6. Interpretation of Meyer's data

If we interpret Meyer's data in the same manner (Figure 1) we shall find that there are mainly two types of dialects: dialects that like Skåne are characterized by an early pulse and a late pulse in the acute and grave accents respectively, and dialects that like Speaker L's have a late pulse in the acute and two pulses in the grave accent.

There is another striking feature about Meyer's curves: whatever the shape of the acute pitch curve the latter part of the grave curve is similar to it.

I can say with a certain amount of truthfulness: Give me your acute accent and I'll tell you what your grave is like. To perform this trick I do as follows: I squeeze in the acute accent in the latter part of the grave word. Then I extrapolate the curve to the earlier part of the word. Naturally a certain experience is needed to do it in the right way. I have to keep within a certain pitch range and I must follow a certain rule. The rule is that an early pulse in the acute accent brings on a late pulse in the grave one whereas a late pulse in the acute accent gives two pulses in the grave accent. Now we must ask: What is the linguistic implication of this procedure? I think it implies that in the acute accent we do not have a real word tone, only a sentence intonation pulse and it is this sentence intonation pulse that recurs in the second syllable of the grave accent.

The idea that only the grave accent is a real word tone and the acute merely the manifestation of stress is not a new one. It was put forward in general terms by Henry Sweet as early as 1878. A similar position was taken by

Trubetzkoy (1939) who regarded the opposition between the two accents as a privative one, that is, one member of the opposition - the grave accent - is marked by a feature that is lacking in the other - the acute accent. Several other scholars have discussed the problem from various angles and have arrived at the same conclusion (Haugen & Joos 1956, 1962, Haugen 1963, 1967a, Malmberg and Elert 1963, Rischel 1963. For a summary of the discussion see Elert 1970).

According to this general analysis, Swedish intonation should be decomposable into sentence intonation and word intonation for the grave accent only. Öhman's model does not satisfy this requirement since he uses a negative pulse for both accents.

His use of negative pulses is open to many objections. We have seen earlier that it has no immediate physiological basis. It can be criticized also on other grounds. In the Stockholm acute accent, Figure 9 (Öhman, _op. cit._ Figure II-B-4), a negative pulse is superimposed on a sentence intonation step in order to produce a rather small dip in the pitch curve which may very well be caused by the oral closure for the initial consonant, in this case _m_. A look at the more schematic Figure 10 (Öhman _op. cit._ Figure II-B-8) reveals that, apart from minor effects, the main function of the negative pulse in Stockholm acute and Malmö grave is to delay the sentence intonation step. In view of these and earlier observations it seems more natural to assume that there is in Swedish just one tone, the grave accent. This word tone is manifested in all the dialects as a delayed sentence intonation pulse which in some dialects (for instance Stockholm but not Malmö) is preceded by an additional positive pulse.

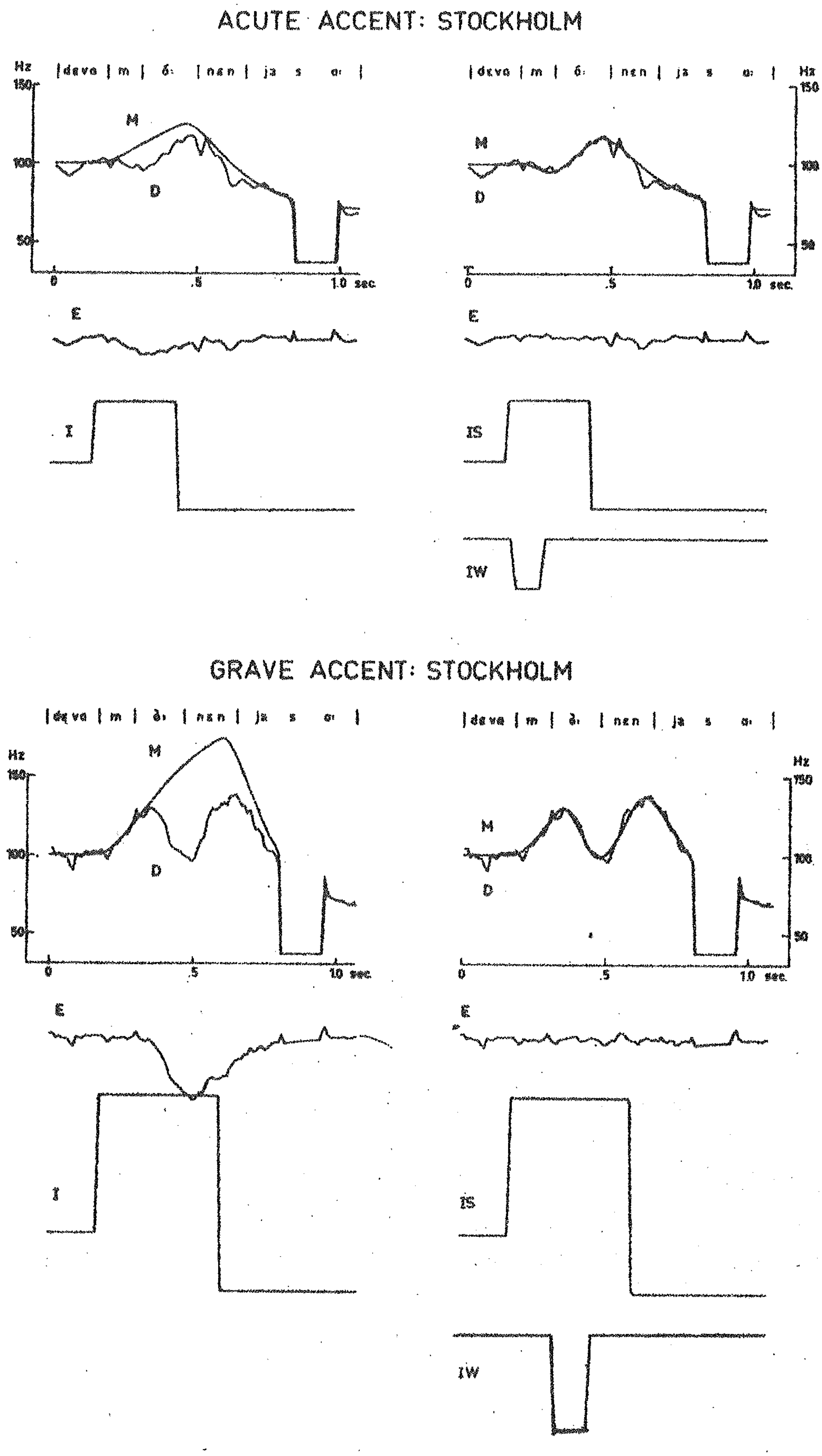


Figure 9.

Fig. II-B-4. Comparison of Stockholm accent patterns with curves calculated by means of intonation model. The pulses marked I, IS, and IW represent model outputs with the same input commands that were used to match the empirical data but with the model constants α and β both set to 1000.

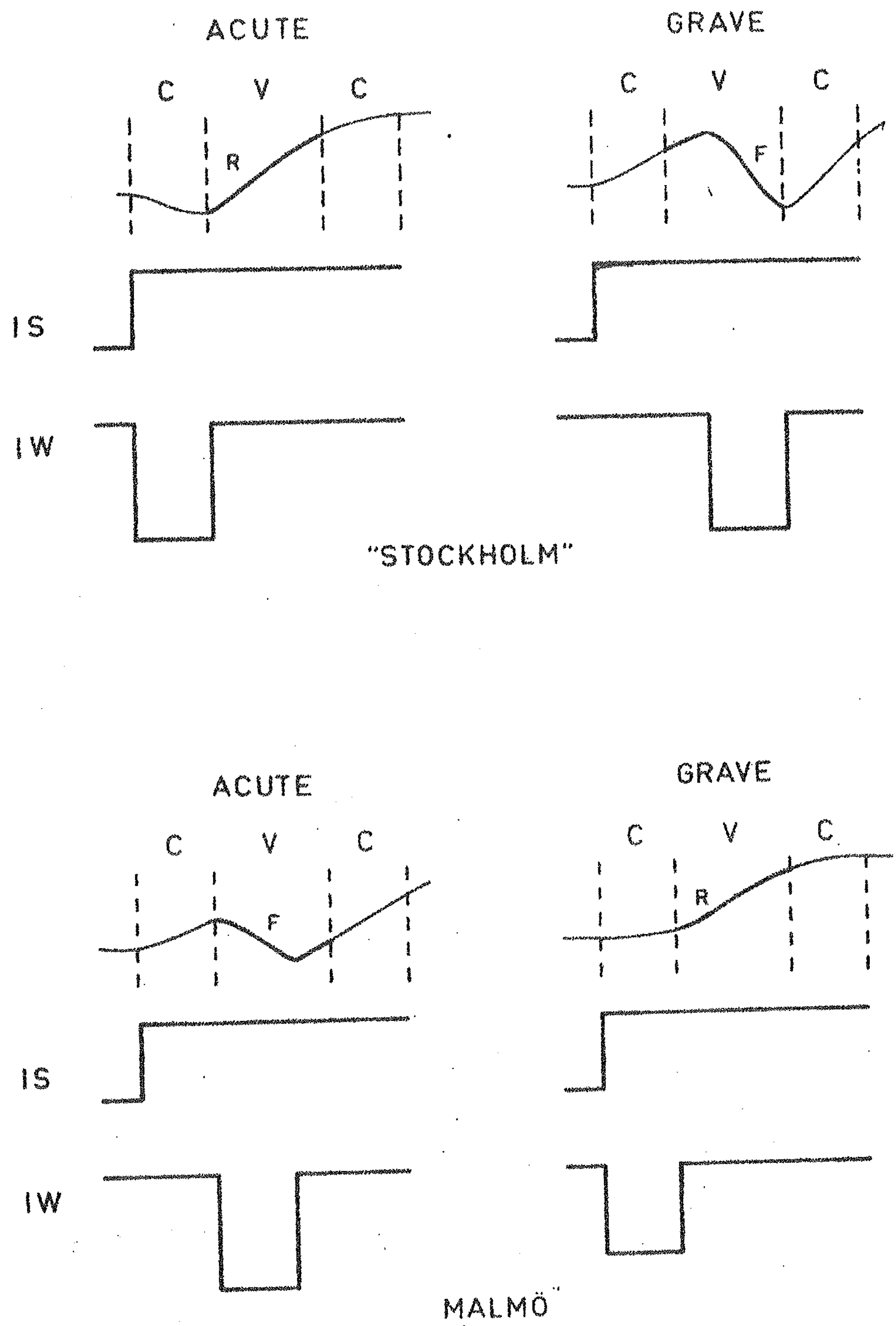


Figure 10.

Fig. II-B-8. Input commands suitable for Stockholm and Malmö accents. The pitch contour has been drawn with thick lines in the vowel segments.

7. Word tones and stress correlates

The dictionary of the Academy marks the accent contrast with numbers that indicates the difference in stress patterns that was supposed to be an invariable characteristic of these words quite independent of the dialectal variation in pitch. The acute accent is marked 4-0, and the grave 3-2, for example $\overset{4}{\text{an}}\overset{0}{\text{den}}$, $\overset{3}{\text{an}}\overset{2}{\text{den}}$. This marking system was introduced well before the time of acoustic measurements. Later, when such measurements could be made it was at first somewhat surprising that no consistent difference in the energy output of the acute and grave accented words could be found.[1]

Today it is well known that subjective impressions of stress are related to the total speech effort and only partly determined by the energy of the speech wave.[2] Increased pulmonary effort in connection with stressed syllables was demonstrated in EMG recordings by Ladefoged, Draper, and Whitteridge (1958), increased laryngeal effort is evident from Ohala's EMG work (Ohala 1970) and Harris and co-workers (Harris *et al*. 1968) showed that increased activity in some articulatory muscles is a constant correlate to stressed syllables.

Our EMG data may to a certain extent explain the difference in subjective stress attributed to the two accent patterns. Look again at Figures 5 and 8. For both speakers - notwithstanding the dialectal difference in their pitch curves - the second syllable of the grave words coincides with more activity in the vocalis and cricothyroid muscles than the corresponding syllable of the acute words. The second syllable of the grave accented words then seems to require a greater effort from these muscles.

---

1 In additional data, that I have received later, the averaged audio signals from Speaker E have a higher amplitude in the second syllable of the grave accented words than in the acute ones.

2 For a clarifying survey of the concept of stress and its acoustic and physiological correlates, see Lehiste, 1970.

The subjectively felt difference in stress connected with the acute and grave accents is perhaps a reflex of the difference in laryngeal activity that may be common to most dialects.

Acknowledgement

I wish to thank the Institute of Logopedics and Phoniatrics, Tokyo, for inspiring interest, skilful cooperation and generous help.

References

Arnold, G.E., Physiology and pathology of the cricothyroid muscle, Laryngoscope 71, pp. 687-753, 1961

Buchthal, F., Electromyography of intrinsic laryngeal muscles, Quaterly Journal of Experimental Physiology, Vol. XLIV, 1959

Buchthal, F. and Faaborg-Andersen, K., Electromyography of Laryngeal and Respiratory Muscles, Annals of Otology, Rhinology and Laryngology, Vol. 73, 1964

Elert, C.-C., Ljud och ord i svenskan, Stockholm, 1970

Elert, C.-C., Phonologic Studies of Quantity in Swedish, Stockholm, 1964

Faaborg-Andersen, K., Electromyographic investigation of intrinsic laryngeal muscles in humans, Copenhagen, 1957

Faaborg-Andersen, K., Electromyography of Laryngeal Muscles in Humans. Technics and Results, New York, 1965

Gårding, E., Fujimura, O., and Hirose, H., Laryngeal control of word tones, Annual Bulletin, No. 4. Research Institute of Logopedics and Phoniatrics. University of Tokyo, 1970

Harris, S., Gay, T., Sholes, G.N., and Lieberman, P., Some Stress Effects on Electromyographic Measures of Consonant Articulations, Haskins Laboratories, SR 13/14, pp. 137-151, 1968

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

Haugen, E., On the rules of Norwegian tonality, Language 43, pp. 185-202, 1967

Hirano, M., Koike, , and von Leden, , The Sternohyoid Muscle during Phonation, Acta Oto-Laryngologica 64, pp. 500-507, 1967

Hirose, H., Forthcoming report. Annual Bulletin. No. 4. Research Institute of Logopedics and Phoniatrics. University of Tokyo, 1970

Ladefoged, P., Draper, M.H., and Whitteridge, D., Syllables and Stress, Miscellanea Phonetica 3, pp. 1-14, 1958

Lehiste, I., Suprasegmentals, Cambridge, Mass and London, 1970

Malmberg, B., Analyse instrumentale et structurale des faits d´ac ents. Proceedings of the Fourth International Congress of Phonetic Sciences, pp. 456-475, 1962

Meyer, E.A., Die Intonation im Schwedischen, Teil I, Stockholm, 1937

Meyer, E.A., Die Intonation im Schwedischen, Teil II, Stockholm, 1954

Ohala, J.J., Aspects of the control and preduction of speech, Working papers in phonetics 15, Los Angeles, 1970

Ohala, J.J., Hirano, M., and Vennard, W., An electromyographic study of laryngeal activity in speech and singing, Report of the 6th International Congress on Acoustics, B-1-2, Tokyo, 1968

Ohala, J.J. and Hirose, H., An electromyographic study of laryngeal activity in speech and singing, Fall meeting, Acoust. Soc. Japan, 1969

Rischel, J., Morphemic tone and word tone in Eastern Norwegian. Phonetica 10, pp. 154-164, 1963

Smith, S., Stødet i dansk rigssprog, København, 1944 a

Smith, S., The Danish Stød, Summary and curves, Copenhagen, 1944 b

Sonesson, B., The functional anatomy of the speech organs, Manual of Phonetics, ed. Malmberg, pp. 45-75, 1968

Simada et al., Forthcoming report. Annual Bulletin. No. 4. Research Institute of Logopedics and Phoniatrics. University of Tokyo, 1970

Sweet, H., Sounds and forms of spoken Swedish, Transactions of the Philological Society, pp. 478-543, 1878

Wessén, E., Svensk språkhistoria, Lund, 1968

Öhman, S.E.G., Word and sentence intonation: a quantitative model. Speech Transmission Laboratory, QPSR, 2-3, pp. 20-54, 1967 a

Öhman, S.E.G., Mårtensson, A., Leanderson, R., and Persson, A., Crico-thyroid and vocalis muscle activity in the production of Swedish tonal accents: a pilot study, Speech Transmission Laboratory, QPSR, 2-3, pp. 55-57, 1967 b.

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 ànden (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 búren (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.

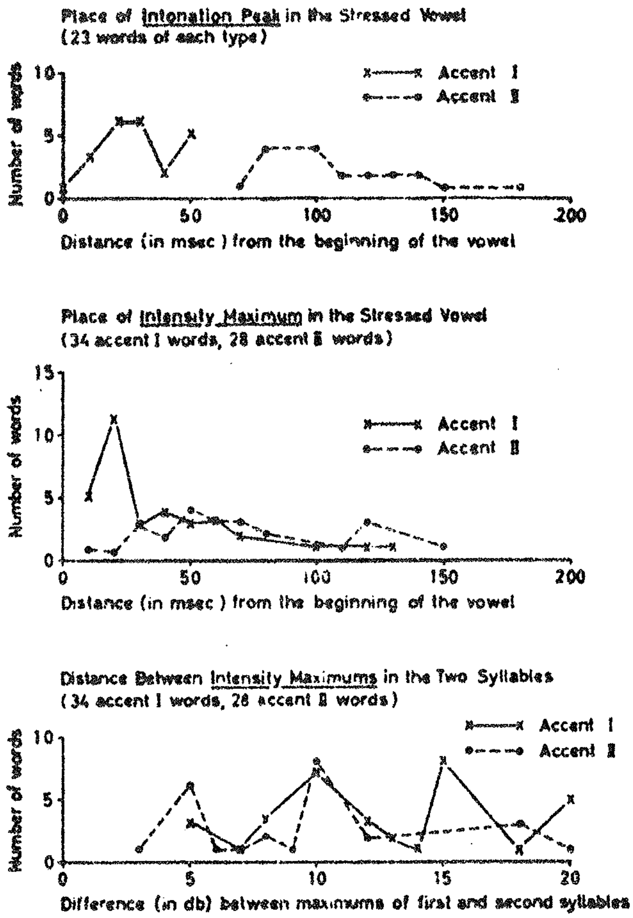


Fig. 1. Data on frequency and intensity overlapping in Swedish accent-1 and accent-2 words (Malmberg, 1962).

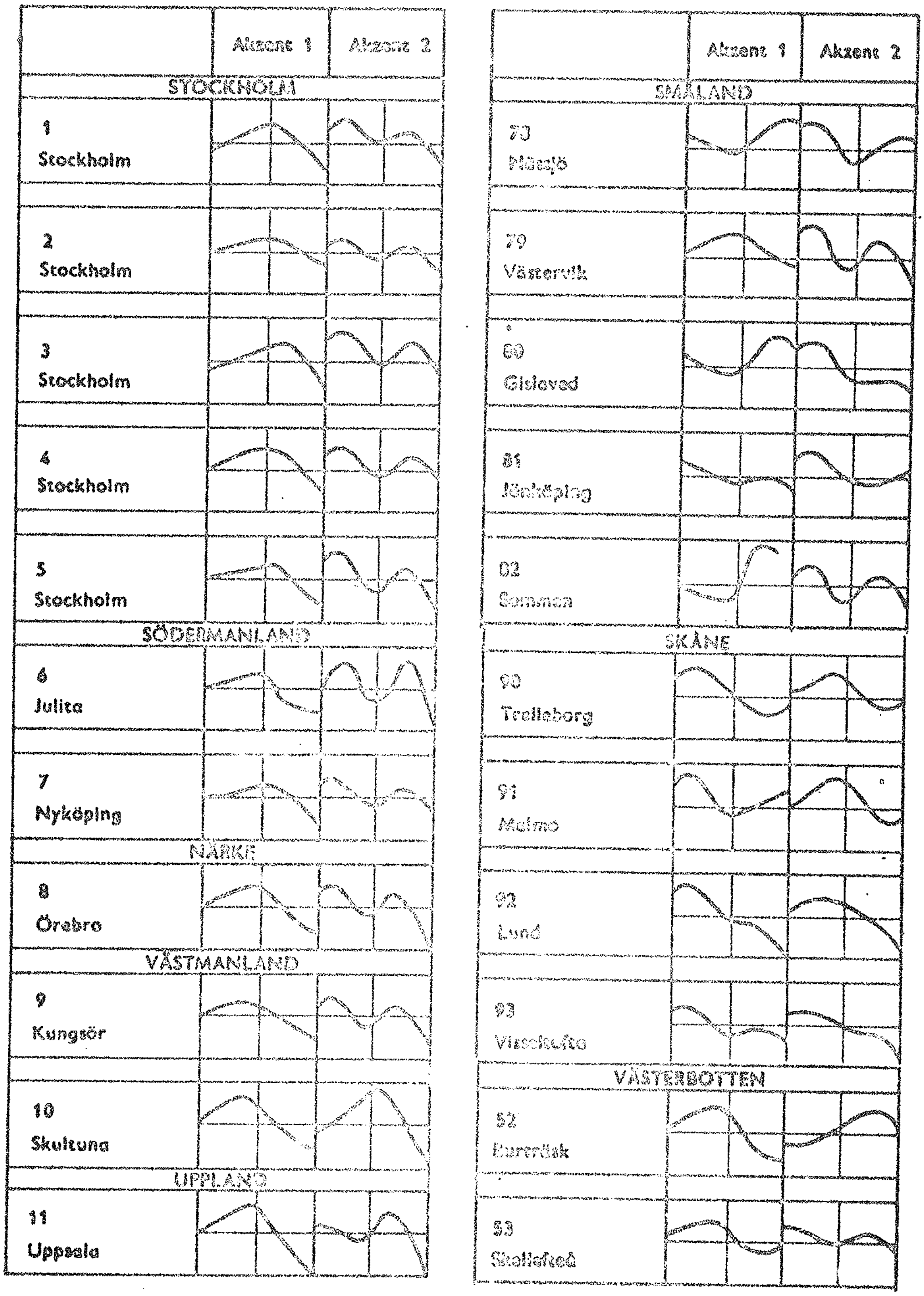


Fig. 2. Some frequency curves from Meyer (1954).

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) våken (the hole in the ice) : vàken (awake)

2) skótten (the shots) : skòtten (the Scot)

3) tómten (the [building] grounds) : tòmten (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.

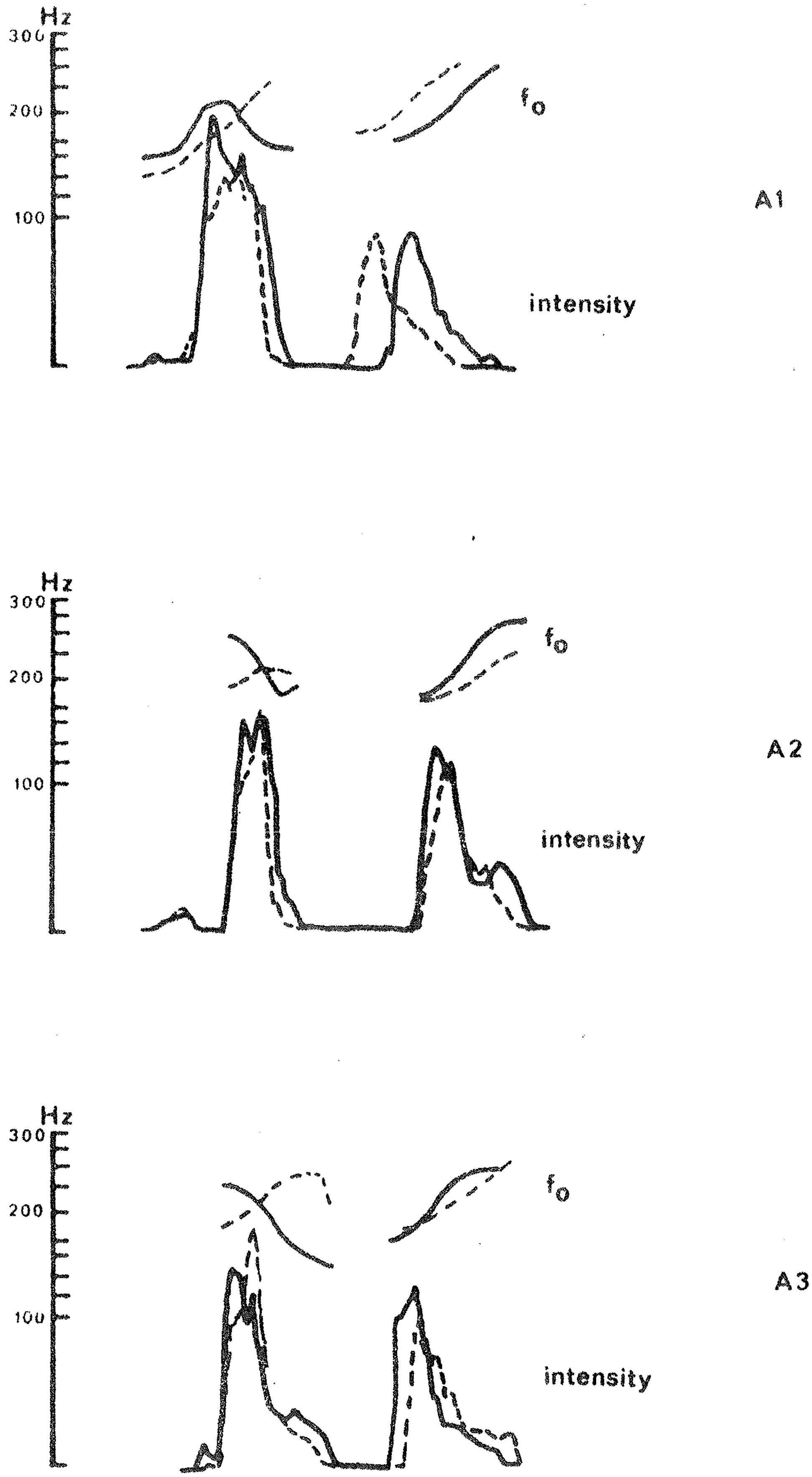


Fig. 3. Fo and intensity curves for speaker A's váken:vàken (A1), skótten:skòtten (A2), and tómten:tòmten (A3). Broken lines = grave words.

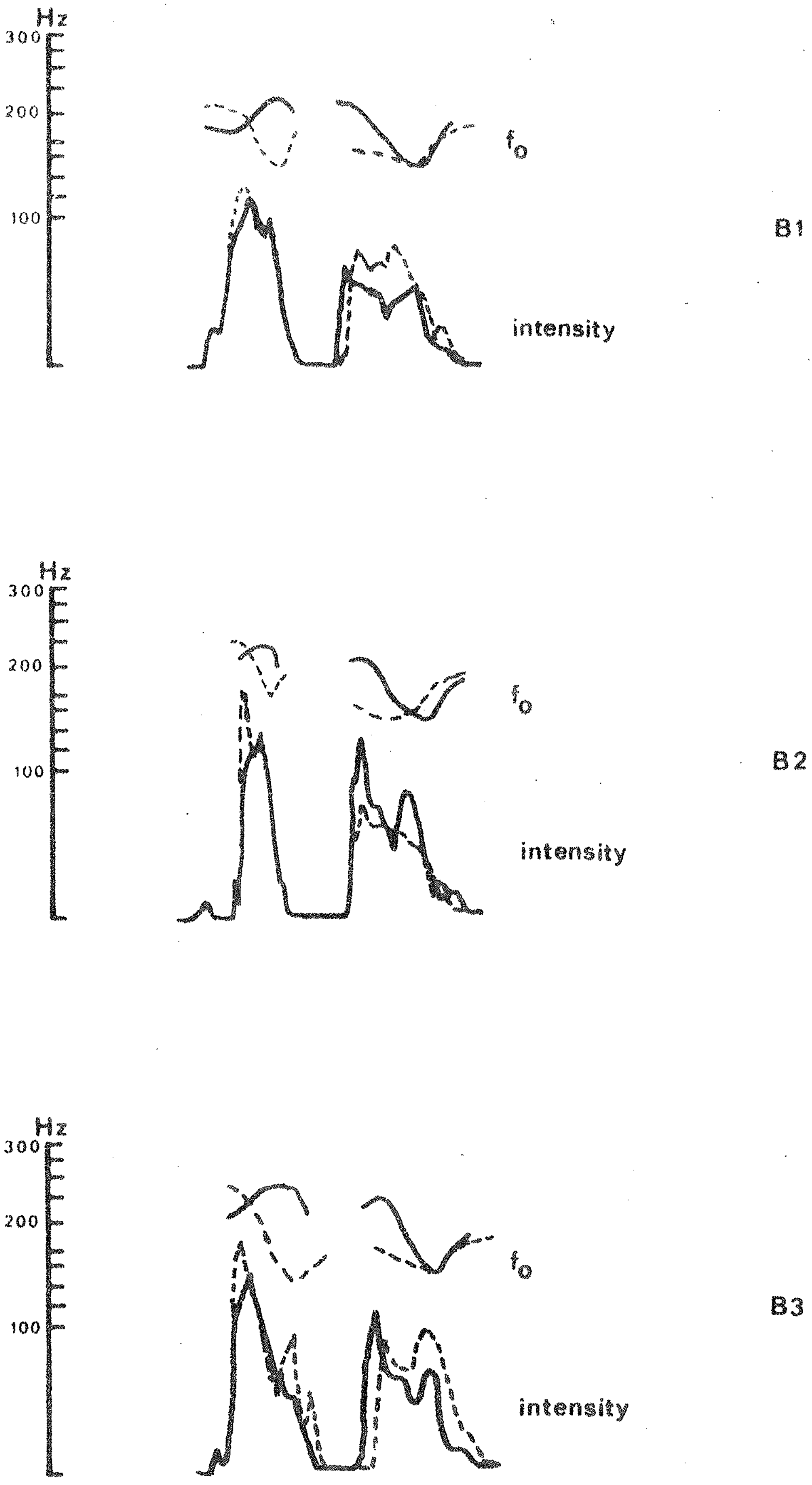


Fig. 4. Fo and intensity curves for speaker B's <u>váken:vàken</u> (B1), <u>skótten:skòtten</u> (B2), and <u>tómten:tòmten</u> (B3). Broken lines = grave words.

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 $+3\sigma$ from the means, for vàken by $+2\sigma$. They are obviously fairly sure that grave words had been produced. This is also the case for tômten ($+3\sigma$), 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 $\pm n\sigma$. E.g. $+3\sigma$ 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. $\pm 2\sigma$ 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 $+3\sigma$, 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_o 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
váken	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
váken	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 $+3\sigma$: for all West Bothnian stimuli, and for Scanian skôtten and skòtten. Also tômten for this dialect indicates (relative) certainty ($+2\sigma$). 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 %	65 %	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 $+3\sigma$ (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 $+2\sigma$ (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
vá-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 váken 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.

vá(-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 %

và(-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).

vá(-ken)

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

và(-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 $+3\sigma$ in all cases.

For speaker B (fig. 14) the deviation for váken 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 vàken instead. Then we get a dramatic shift to positive deviations ($+3\sigma$ for stimuli No. 8 and 10, while No. 9 only gives $+1\sigma$, i.e. guessing).

våken (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 %
Broup 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 $+3\sigma$. 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 ($+2\sigma$).

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 våken.[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 våken, 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 våken 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 (våken, 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 våken 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	vǻ(ken)	290 msec	30	–	70	–	–	–	50	–	–	–	–	–
A	vå̀(ken)	270 msec	55	–	155[1]	–	–	–	135[1]	–	–	–	–	–
B	vǻ(ken)	230 msec	100	100	120	–	80(40)[2]	80(40)[2]	100	–	60(20)[2]	60(20)[2]	60	–
B	vå̀(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

– (1955), 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.