Pitch as a correlate of Swedish word accent

Jassem, W.

journal: STL-QPSR
volume: 3
number: 4
year: 1962
pages: 004-014

http://www.speech.kth.se/qpsr
B. PITCH AS A CORRELATE OF SWEDISH WORD ACCENT

The vocabulary of the Swedish language is known to include word pairs which, whilst identical on the phonemic level, differ only by a suprasegmental feature referred to as "word accent". Examples are 'buren "the cage", vs. 'buren "carried", past part., 'tanken "the tank" vs. 'tanken "the thought" etc. Although a count of the number of such pairs might be interesting in itself*, it would not reflect what is usually referred to as the "semantic load of an opposition", because the difference in the frequency of occurrence between the numbers of such pairs may either be large or small. The dynamics of linguistic form further complicates this problem. Many pairs may be found in which one member is the plural imperative, which is entirely obsolete e.g. 'köpen "the purchases" 'köpen "purchase" (plur.imper.). Even a pair like 'överste "uppermost" vs. 'överste "colonel" might be disputable, the former now occurring in written texts only. In the present writer's opinion the existence of such "minimal pairs" is not a necessary condition for the establishment of relevant (distinctive) word accent. In other words, the fact that such pairs do exist is incidental (though it may be convenient for some experimental work), and the relevance of the accent may be proved regardless of them. It is sufficient to show that, 1. the distinction is not free, and 2. that the distinction is not predictable. This proof will here be given in a brief outline.

1. The distinction is not free because word (and utterance) forming sequences of phonemes can be found which (under given conditions of stress) are only combined with either one or the other accent. There is, for instance, 'vinkel, but no 'vinkel, 'gammal, but no 'gammal, 'neger, but no 'neger, etc.

2. That the distinction is unpredictable can be shown as follows: Let us compare, for instance, 'nyckel with 'cykel

*According to a private communication from dr. C.-C. Elert, Stockholm University, up to 500 pairs can be found.
[ˈsʌkə] These words differ by accent and also by the initial consonant. The accent is, however, independent of this, because the words ‘neve’ and ‘seve’ show the same phonemic difference, whilst both having accent 1. By extending this line of reasoning the relevance of the Acc. 1 vs. Acc. 2 distinction may be proved by reference to words with entirely different phonemic structure and applied to words taken in isolation (i.e. constituting complete utterances) as well as to words considered in wider contexts.

The so-called melodic word accent in Southern Swedish has received much more attention than that used in Stockholm. In an early study E. Malmberg (4) suggested that the essential difference is that between a single pitch peak (Acc. 1, "acute") and a succession of peaks (Acc. 2, "grave"). In a later study he modified his description, establishing an opposition between a sharply falling ("starkt fallande") Acc. 1 and a slightly rising ("svagt stigande") Acc. 2 for Southern Swedish (essentially the dialect of Scania) (7). He also found that the peak of the pitch curve is nearer the beginning of the stressed syllable in Acc. 1 words than it is in Acc. 2 words (6). The only study based on any larger-scale measurements on the melodic accent in Stockholm is still that by E.H. Meyer (8), who also shows the drastic differences in the "realizations" of the distinction in the different dialects. Essentially, he arrives at a rise-fall for Acc. 1 and a fall-rise-fall for Acc. 2 in Stockholm Swedish. He also points out differences in the tone preceding the stressed syllable ("upptakt"), this being lower before Acc. 2 and higher before Acc. 1. A very important study of the word accents in a Scandinavian language, methodologically as well as theoretically, is that by E. Haugen and M. Joos (3). The most striking point in this treatment of the subject is that an utterance is divided into measures - purely phonetic units delimited by stresses - rather than words, and the two accents are attached to those measures and integrated into a description of overall intonation in terms of (tentative) pitch levels. The most extensive study of pitch in Swedish is that by K. Hadding-Koch (2). Like Haugen and Joos she deals with measures and pitch levels, but her measures are delimited by junctures, and her pitch levels are independent of word accent. K. Hadding-Koch was able to show
that word-accent distinctions, which previously had been studied either in isolated words or in very restricted contexts, are essentially preserved in connected speech. Again, Southern Swedish was the subject of this study. C. Witting \((10)\) suggests that 'sentence intonation' tends to obliterate word accent differences (p. 208).

The main aim of the present investigation is to collect and interpret a body of data on the pitch pattern in connected speech as related to the word accents in the variety of Swedish spoken by educated residents of Stockholm, the material being connected speech. Units of word length have been chosen because a division of utterances into measures (or possibly other units) would either involve subjective judgements (as in refs. (3) and (2)), which the author decided to avoid on principle, or else necessitate a preliminary investigation which would make it impossible to attack the main problem within the available time. It seems that the results which will here be presented show that working with units of word length leads to generalizations that are no less consistent than those based on measures.

The material on which the following analysis is based consists of the first 243 words of S. Dagerman's short story "Att döda ett barn", read twice by each of the two subjects, BL and SÖ, and recorded on tape in an anechoic chamber. The method of data collection was a combination of perceptual analysis and measurement of the speech wave.\(^*\) The perceptual analysis was performed with the aid of a tape scanner with a rotating playback head for repeated reproduction. A section of the recording could be gated out and the duration of the gated section finely adjusted within the range of interest. For measurements of the speech wave ink-written oscillo-grams at 500 mm/sec and narrow-band spectrograms with a scale of

\[^*\]"It must be recognized that from a linguistic point of view the impressions registered by the ear alone are of highest importance, and cannot be eliminated in favour of mechanical recordings, no matter how perfect", see \((3)\) , p. 42). It might be added that auditive impressions are also an important subject of study for various technical problems of speech transmission.
40 mm/kc were used. The oscillograms were read with the aid of a magnifying glass, the accuracy being \( \pm 0.1 \) mm. The fourth harmonic was traced on the spectrograms giving an accuracy of approx. \( \pm 3 \) c/s. Both oscillograms and spectrograms were made so that measurements could be taken from whichever of the two appeared more reliable, depending chiefly on harmonic composition and fundamental frequency. The primary aim of the perceptual analysis was to establish, (a) the type of tone within syllables (i.e. level, falling, or rising syllabic pitch)*, and (b) the pitch relations between successive syllables (i.e. the relations between the final pitch of one syllable and the initial pitch of a following syllable). The assumption that the syllable is the natural substructural pitch segment in speech perception has generally been assumed in phonetics and was corroborated in Root's study (9). A curve representing \( F_0 \) (the fundamental frequency) as a function time is only broken at pauses and intervals corresponding to voiceless consonants. It can be segmented into syllables, however, at phonemic border points which are clearly determinable on spectrograms. A consonant which does not phonetically belong to any syllable (as in VCV combinations) has been disregarded. The type of tone in every syllable and the relation between successive syllables were first established with the help of the tape-scanner. Whilst the auditive syllabic tones could in every case be specified unambiguously as belonging to any of the above-mentioned types, the registered wave naturally showed an almost infinite variety in the type of \( F_0 \) time function within the syllable. In order to correlate this function with the perceived units, each syllable as registered by the spectrograph or the oscillograph was classified as belonging to one or the other of the same types of syllabic tones as those perceived, i.e. as even, falling or rising according to the observable predominant shape. The following general differences between the perceived syllabic tones and the corresponding \( F_0 \) time function segments were observed: 1) Short-time sudden changes of fundamental frequency including only a few periods and mostly appearing at the transitions between consonants and vowels were not perceived. 2) The

* Composite (e.g. falling-rising) syllabic tones were not found in the material.
F₀ time function in a consonant which only showed between 1 and 3 harmonics on the spectrograms had no or very little influence on the perceived pitch. 3) Falls and rises in the F₀ time function contained within relatively short syllables were perceived as even tones.* The effect of duration on the perception of syllabic tones was studied on a material consisting of the first 124 syllables of the first reading of the text (BL). The length of the voiced syllabic core was measured and the type of predominant F₀ time function was compared with the type of the perceived tone type. In 81 syllables the predominant F₀ time function was non-constant. These were divided into 5 groups according to duration and the following table shows the percentage of syllables in each group perceived as not level:

<table>
<thead>
<tr>
<th>Duration in msec</th>
<th>Percent perceived as not level</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>45 - 80</td>
<td>18</td>
</tr>
<tr>
<td>85 - 120</td>
<td>40</td>
</tr>
<tr>
<td>125 - 160</td>
<td>58</td>
</tr>
<tr>
<td>165 - 211</td>
<td>82</td>
</tr>
<tr>
<td>200</td>
<td>88</td>
</tr>
</tbody>
</table>

The perceived syllabic tones and intersyllabic relations were compared with the corresponding spectrographic and/or oscillographic data, and a discrepancy which could not be ascribed to the 3 general factors named above was found in approx. 3 per cent of the cases. These were subjected to a renewed perceptual examination and now found to be in agreement with the measured values. Thus the total gross data eventually showed no discrepancy between perception and measurement apart from differences which are being assumed as inherent (i.e. not due to subjective error in perception).

The analysis of the gross data was limited to words of two or more syllables, there being no contrast of word accent in monosyllables. Each word in BL's first reading was compared with the corresponding word in his second reading and the same procedure

* Cf. the detailed findings in ref. (9) and also ref. (2), p. 58-59.
was applied to SÖ's two readings. For each of the two speakers, therefore, each text-word received two descriptions of the pitch pattern. There were 36 two-syllable Acc. 1 words in the text and their patterns are summarized in Table I - 1a and I - 1 b. The symbols representing the pitch patterns are to be interpreted follows: (A) The letters refer to syllabic tones; L = level, F = falling, R = rising. (B) of both syllables are level a minus means "second syllable lower" a plus "second syllable higher" and no sign means "the same pitch". F or R is followed by a minus or plus sign only if there is a shift of pitch in the opposite direction. Thus F+L means "first syllable falling, second level but higher than the end of the fall". FL is to be read "fall followed by level not higher than the end of the fall". The tabulations do not indicate in which of the two readings a particular pattern occurred.

| Table I - 1a |
| Pitch patterns in Acc. 1 words in the two readings by BL |
| LL | L+L | L-L | FL | R-L | F+L | L-L | FR | LL | LL | LL | L+L | L+L | L-L | L+L |
| LL | L+L | L-L | FL | R-L | F+L | FL | F+L | FL | L+L | L-L | RL | FR | FF | L+L | F+L |
| 2 | 10 | 1 | 2 | 1 | 1 | 3 | 1 | 1 | 1 | 3 | 1 | 1 | 2 | 1 | 2 | 3 |

| Table I - 1b*** |
| Pitch patterns in Acc. 1 words in the two readings by SÖ. |
| Pitch patterns |
| LL | L+L | FL | RL | LL | LL | L+L | L-L | L-L |
| LL | L+L | FL | RL | L+L | L-L | RL | R-L | L+L |
| No. of text-words |
| 5 | 13 | 2 | 1 | 6 | 3 | 2 | 1 | 2 |

It can be seen from the tables that the most common pitch pattern of Acc. 1 words is L+L which occurs consistently (i.e. in

* No distinction was made according to the relation between the end of the fall and the beginning of the rise since in both cases there was a change of direction.

** There was a single case of a fall followed by another lower fall.

*** The number of text words is 36 in BL and 35 in SÖ because SÖ read the word ['kamera] once with three syllables and then with two, so this word is not included in the table.
the same text-words) in both readings 10 times in Table I - 1a
and 13 times in Table I - 1b, with occasional variations with other
patterns. A general rising word intonation is sometimes also
represented by RL in both tables, and there are a few specimens of
R+L. A striking feature of BL's Acc. 1 word intonation is the not
infrequent occurrence of a fall-rise. Both speakers also used falling
word intonations: FL or L-L. Altogether patterns containing some
form of rise in pitch within the word totalled 62 per cent in BL and
59 per cent in SÖ. Falling word intonation (L-L or FL) occurred in
all four readings mostly in utterance-final positions.

Table I - 2a
Pitch patterns of Acc. 2 words with the stressed syllable followed
by one stressed syllable in the two readings of BL.

<table>
<thead>
<tr>
<th>Pitch patterns</th>
<th>L-L</th>
<th>FL</th>
<th>FR</th>
<th>F+L</th>
<th>L-L</th>
<th>FL</th>
<th>F+L</th>
<th>L-L</th>
<th>LL</th>
<th>L+L</th>
<th>LL</th>
<th>L+L</th>
<th>LL</th>
<th>F+L</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of text-words</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table I - 2b
Pitch patterns in Acc. 2 words with the stressed syllable followed
by one unstressed syllable in the two readings of SÖ.

<table>
<thead>
<tr>
<th>Pitch patterns</th>
<th>L-L</th>
<th>FL</th>
<th>FR</th>
<th>F+L</th>
<th>L-L</th>
<th>F+L</th>
<th>F+L</th>
<th>L+L</th>
<th>LL</th>
<th>L+L</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of text-words</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Owing to slight differences in pronunciation BL had 11
text words with more than one syllable following the stress, and SÖ
12. There were considerable variations in the syllabic patterns of
these polysyllabic words. In view of the relatively small number
of specimens of words with more than one syllable following the
stress it was decided to categorize them according to the general
word intonation. Thus for instance F+L+L, FRL and F+LL were pooled
together as falling-rising word intonations. A comparison of the

*There was one case of a fall followed by another fall with change
of direction (this is a falling-rising-falling word intonation).
text-words between the two readings of BL and SÖ gives the following result:

<table>
<thead>
<tr>
<th></th>
<th>BL</th>
<th>Word intonations</th>
</tr>
</thead>
<tbody>
<tr>
<td>fall-rise-fall fall-rise fall-rise-fall</td>
<td>word intonations</td>
<td></td>
</tr>
<tr>
<td>fall-rise-fall fall         fall-rise</td>
<td>No. of text-words</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SÖ</th>
<th>Word intonations</th>
</tr>
</thead>
<tbody>
<tr>
<td>fall-rise-fall fall-rise fall-rise-fall</td>
<td>word intonations</td>
<td></td>
</tr>
<tr>
<td>fall-rise-fall fall-rise fall-rise</td>
<td>No. of text-words</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

Thus all specimens of words with more than one syllable following the stress both in BL and SÖ had an intonation beginning with a fall. Altogether 84% of the specimens of Acc. 2 in BL and 83% in SÖ have some form of an initial fall in the word intonation.

The majority of Acc. 1 words in the whole corpus, then, have a rising intonation (sometimes preceded by a fall) whilst a still greater majority of Acc. 2 words have a falling intonation (not infrequently followed by a rise, and in words with more than one syllable following the stress, by a rise and another fall).*

All the rising inflections in Acc. 1 words and all the falling inflections in Acc. 2 words were measured by establishing the frequency of $F_0$ at the beginning and the end of these inflections. Fig. 1 - 2a represents the distribution of $F_0$ frequency of a rising inflection in Acc. 1 words and Fig. 1 - 2b the distribution of $F_0$ frequency at the end of the rise, in BL's both readings. It will be seen that the initial frequency of BL's rising inflection in Acc. 1 words varies between 85 and 145 c/s and the final frequency between 91 and 190 c/s, with a few not very marked peaks in both

*It would seem that Meyer's description of Acc. 1 as a rise-fall and Acc. 2 as a fall-rise-fall, which has almost become a textbook standard, applies to a form of Stockholm speech is at least somewhat emphatic. These two intonations would also appear to be limited to an utterance-final position, which includes word spoken in isolation. Meyer himself observed deviations from these "standard" forms in connected speech, but his study of the word-accents in longer utterances was treated as a side-issue, see ref. (8), pp. 61-65.
Fig. I - 2. Acc 1 words - Speaker BL

Initial $F_0$ frequency of the rise

Final $F_0$ frequency of the rise

Ratio of initial to final $F_0$ frequency
histograms. Fig. 1 - 2 c represents a distribution of the ratios of initial to final frequency in each rise and the values here cluster densely around approx. 0.90. Corresponding distributions histograms for SÖ are represented in Fig. 1 - 3a,b,c. The spread of initial and final frequencies is here smaller than in BL and the peak in the distribution of the ratios lies somewhat higher (approx. 0.95).

Similarly, the frequency at the beginning and the end of the (initial) fall in Acc. 2 words was measured and the ratios for each cases were calculated. Fig. 1 - 4a,b,c shows distribution histograms of the values obtained for BL and SÖ. Whilst the distribution of the initial and final values as well as the ratios appear to be essentially one-peak for SÖ, only the distribution of the final F₀ frequency seems to be of the one-peak type for BL, whilst that of the initial frequency here tends to show three peaks and that of the ratios is rather irregular.*

The intonation of the word itself is, however, not the only distinguishing feature between the two word accents. In Acc. 2 words standing in a non-initial position in an utterance the initial tone of the lexically stressed syllable has been found in BL to be preceded in 92 % of the cases by a syllable with a lower final pitch. The corresponding figure for SÖ is 88 %. In non-initial Acc. 1 words the beginning of the lexically stressed syllable is preceded by a higher tone in 73 % of the cases both in BL and SÖ (with 15 % - BL - and 14 % - SÖ - with the same tone and 12 % - BL - and 13 % - SÖ - with a lower tone). It is noteworthy that in cases where an Acc. 1 word had a falling or falling-rising intonation its lexically stressed syllable was never preceded by a lower tone.

The results of the analysis of the prose passage can therefore be summarized as follows: In an utterance-initial position a distinction is made between words which retain their

---

*The distribution histograms appear to confirm a subjective impression that SÖ tends to speak in less lively tones than BL.
Initial $F_0$ frequency of the rise

Final $F_0$ frequency of the rise

Ratio of initial to final $F_0$ frequency

Fig. I - 3. Acc 1 words - Speaker S6
**Fig. 1 - 4. ACC 2 words - Speaker BL**

**a)**

Initial $F_0$ frequency of the fall

**b)**

Final $F_0$ frequency of the fall

**c)**

Ratio of initial to final $F_0$ frequency
Initial $F_0$ frequency of the fall

Final $F_0$ frequency of the fall

Fig. I - 5. Acc 2 words - Speaker 80
lexical Accent 2 and words which either lose this Accent* or have a lexical Accent 1. If a word retains its lexical Accent 2 it has a falling intonation, at least initially, (counting from the lexical stress) and this fall is preceded by a lower tone.

The material being too small for anything more than tentative assumptions concerning positions other than internal, an experiment was divised to obtain some more information. One crucial problem seemed to be to find whether there is an accent distinction in final positions where both types of words - according to the analysis of the text - may have a falling intonation.

Eight sentences spoken by a native of Stockholm (female voice) were recorded on tape. They all began with the words: Jag använder ... and 4 of them finished with ... som täcken, the other 4 with ... som tecken. Fig. 1 - 3 shows the intonations of the final two words as traced from spectrograms (fourth harmonic). The four cases of /ték:en/ and the four cases of /tö:ken/ were spliced out of a copy of the original tape and presented in random order to 10 listeners (all residents of Stockholm). It can be seen from Fig. 1 - 6 that the essential (and probably the only perceptible) difference in pitch is that the first syllable of täcken is quite consistently higher than that of tecken. 7 out of 10 listeners identified all words correctly. 1 person made 1 mistake, 1 took all to be técken, and the responses of 1 person were quite random. From another copy the last two words of the eight sentences were spliced out. These were presented (in a different order) to the three listeners whose responses had not been correct. The listener who had taken all words to be tecken, now made all his responses correct. So did the one who had made one

---

* Under what conditions word lose their lexical Acc. 2 in an utterance have not been sufficiently investigated. One condition seems to be some kind of "sentence stress" or prominence. In our materials word like "mellan, donna, sin, medan" and verbs followed by short adverbs (e.g. ... "tittar in") have both in BL's and SÖ's readings no fall. Cf, ref.(5),p. 107: Förutsättningen för att någon accentueringsmotsättning skall kunna komma i fråga i svenska, är alltså, att .... ordet i fråga står på tryckstark plats i satsen." However, there are also words in our readings which either have no fall or, being in an internal position, are not preceded by a lower tone, though they probably cannot be described as having no prominence (eg. ...... den vackra bilden ...., i den första byn Cf. ref. (1), p.52.
Fig. 1 - 6. The words "som täcken" (4 versions) and "som tecken" (4 versions) in the order of the experiment. $F_0$ traced from spectrograms.
mistake. The remaining listener did not improve. It will be seen in Fig. 1-6 that the beginning of tecken is consistently preceded by a higher tone whilst that of täcken just as consistently—by a lower tone. Measurements showed that there were no consistent difference between tecken and täcken either in the duration of the phonetic segments, or vowel formant frequencies or amplitude relations. Although this is a very small-scale experiment it made the following conclusion at least highly probable: In an utterance-final position, where both Acc. 1 and Acc. 2 may have a falling word intonation, the interval relations between the stressed and the following unstressed syllable is an important distinguishing cue, the difference in the preceding tone (lower before Acc. 1 vs. higher in Acc. 1) being an additional distinguishing feature.

W. Jassem

References

(1) Danell, G.: Svensk ljudlära, Stockholm (1926)
(5) Malmberg, B.: Svensk fonetik, Lund 1951
(7) Malmberg, B.: "Sydsvensk ordaccent", Lunds Universitets Årsskrift, N.F. Av. 1, Bd 42, Nr. 2 (1953)