On the coordination of articulatory and phonatory activity in the production of Swedish tonal accents

Öhman, S.

journal: STL-QPSR
volume: 6
number: 2
year: 1965
pages: 014-019

http://www.speech.kth.se/qpsr
In general, non-compound standard Swedish words of Scandinavian origin have stress on the first syllable. The other syllables are unstressed. Furthermore, the stressed syllable has the rising tone if the stem contains exactly one vowel, otherwise it has the falling tone*. The classical examples are ånden (the duck) and ånden (the spirit) the stems of which are and and ande, respectively.

As the names imply the rising tone is ordinarily associated with a rising fundamental frequency ($f_o$) and the falling tone with a falling $f_o$. However, the description of the Swedish tonal accents as being certain combinations of a (binary) stress with one of two tones on a syllable is essentially based on introspective analyses by native speakers of the language and requires separate justification in terms of objective phonetic measurements.

* In the traditional terminology words which have the rising tone are said to have "accent I" or "the acute accent". Words with the falling tone are said to have "accent II", or "the grave accent". Compound words with the falling tone are also said to have accent II. It should be noted, however, that compound words have two stressed syllables (the first syllable of the first and last stems are both stressed) and that the second one of these has the rising tone. Owing to this circumstance one can find pairs of words both of which have accent II in the traditional sense and both of which consist of the same sequence of phonemes but which nevertheless differ prosodically (and semantically). An example is plåskande ("splashing") vs. flåskande (=bottle spirit). (The signs ′ and ″ denote rising and falling tone, resp.) Except for the inessential p/f contrast the only phonetic difference between the two words of this example is that flåskande has stress and rising tone on its second syllable whereas plåskande does not. Although traditional linguists have been aware of this important distinction their accent I/II-terminology fails to reflect it. For this reason we prefer the terms "rising tone", "falling tone", and "(binary) stress".

These facts will be treated in more detail in a paper on generative phonology of the Swedish verb which is presently in preparation.
Question

The purpose of the present report is to present some data on the relation between the $f_0$ contour and the timing of the various articulatory gestures of non-compound words with falling tone. The question asked is whether certain properties of the $f_0$ contour can be found which are invariant under certain changes of the segmental phonemic contents of the words.

Method

The speech material that was studied consisted of bisyllabic nonsense words of the forms $C_1V:C_2:a$ and $C_1VC_2:a$ which were spoken in the sentence frame

\[ /s\text{ej}_{\text{a}}-\text{I}_{\text{t}}\text{h}/ \] (to say _ _ again)

The vowels and consonants used in one group of words were $V: = /\text{a}/$, $V = /\text{a}/$, $C_1 = /\text{g}/$ or $/k/$, $C_2 = /d/ \text{ or } /t/$. In another group of words $C_1$ was $/q/$, $/t/$, $/f/$, or $/v/$ and $V: = /\text{a}/$, $C_2 = /d/$. These words conform to the admissible phonemic patterns. In fact $/g:\text{ta}/$ (=street), $/k:\text{t}:a/ (=\text{she-cat})$ and $/\text{v}:\text{da}/ (=\text{to wade})$ exist. In Swedish stressed syllables a consonant is automatically short after a long vowel and long after a short vowel.

The purpose of the frame was to reduce the influence of the sentence intonation, particularly that of the steeply falling terminal contour, on the tonal accent of the test words. Moreover, the presence of the final vowel in $/s\text{ej}_{\text{a}}$/ made length measurements on initial voiceless stops of the test words possible. All syllables of the frame were unstressed.

Each of the twelve different words were written on five file cards. The deck of cards was then shuffled. A male native speaker read the words from the cards in an anechoic chamber while a tape-recording was made.

Wide-band and narrow-band spectrograms were made of each utterance. The frequency scales used were 20 mm/kc and 40 mm/kc, respectively. Fig. II-B-1 shows an example of these spectrograms for the utterance $/g:\text{da}/$. The fundamental frequency was measured from the tenth partial (traced out with a black line in Fig. II-B-1b) as seen in the narrow-band displays.
Fig. II-B-1a. Sound spectrogram (narrow band) of the utterance "stja glis ijsn." The vertical lines illustrate the segmentation procedure discussed in the text.

1b. Narrow band spectrogram of same utterance with expanded frequency scale. Black line follows the tenth harmonic.

1c. Superimposed tracings of the tenth harmonic from the five readings of the same utterance. Vertical scale shows fundamental frequency.
The wide-band displays were used for segmentation. The following segment boundaries were marked.

**Voiced stops:** the beginning and end of the stop gaps.

**Voiceless stops:** the beginning of the silent interval (last voice spike of preceding vowel), the beginning of the aspirative release segment, and the first voice spike of the following vowel (the latter boundary was marked only in word initial stops. Word medial voiceless stops are unaspirated but usually end with a brief fricative transient which was marked as the final boundary).

**Voiced fricatives:** the points of most rapid decrease and increase of $F_1$.

**Voiceless fricatives:** the last voice spike of the preceding vowel and the first of the following vowel.

The tenth partial was traced on transparent paper and the segment boundaries were marked off. This process was repeated for each of the five readings of each word so that five superimposed tracings of the tenth partial were obtained on the same paper. The segment boundaries of the initial vowel of the test word were used as time references. Finally, a visually determined average of the five superimposed curves for the tenth partial was traced on a different transparent paper. On this paper the median time locations for the various segment boundaries were also marked. Fig. II-B-1c shows as a typical example the five superposed $f_0$ tracings for the word *glida*.

**Results**

The $f_0$ contours of the different words could now be compared with each other. This comparison showed that the general shape of the $f_0$ contours was very much the same in all the words. Certain differences could be observed, however. In particular, the starting frequency of $f_0$ in the stressed vowel was different after different initial consonants. It was highest after the voiceless stops (/k/), lowest after the voiced consonants (/g/, /j/, and /v/) and intermediate after the voiceless fricatives.
(/f/ and /g/). A difference was also visible in the unstressed vowel of the test word; in general, the pitch started from a somewhat lower frequency after the voiced consonants in this vowel.

The lowering of the pitch during, or close to, a voiced oral consonant should be expected on account of the increased pharyngeal pressure automatically induced by the oral closure. The pharyngeal pressure increase will cause the pressure drop across the glottis to decrease, thereby reducing the Bernoulli force and hence also the pitch (1). The difference in pitch after /k/ as compared with that after /f/ or /g/ is more difficult to understand. It is hoped that direct measurements on the tracheal and pharyngeal pressures which are presently being prepared will clarify this question.

By taking averages it was possible to derive a small number of basic pitch contours each of which gave an optimal fit to the individual contours of a certain class of words. Those contours are shown in Fig. II-B-2 (top). The class of words on which the averaging was made in each case is indicated by reference to the initial and medial consonants, which define the class in question. As noted above, it is conceivable that the tension of the vocal cords as a function of time is the same in all these cases and that the differences in the \( f_0 \) contours of Fig. II-B-2 are due to secondary factors of an articulatory origin.

The locations of the spectrographic segment boundaries of the different words in relation to their respective common \( f_0 \) contours are also indicated in Fig. II-B-2 (lower part). Obviously, the possibility of relating the occurrence of the spectrographic segment boundaries to the pitch contours establishes a common time scale for all the words studied. We can thus determine whether a given segment of a certain word occurs earlier or later than an arbitrary segment of any other word in relation to this common time scale. It should be noted that when the segment boundaries are adjusted in this manner the goodness of fit of the averaged \( f_0 \) contours of Fig. II-B-2 to the individual contours is better than \( \pm 5 \) c/s wherever measurable.

The optimal placement of the segment boundaries in relation to the pitch contour brings out an interesting relation concerning the timing of the voiced and voiceless cognates of the
Fig. II-B-2. (top) Derived pitch contours for the whole word material studied. Explanation in text.
(lower part) Durations of spectrographic vowel and consonant segments of the words studied. Cross-hatched areas indicate vowel segments, blank areas represent stop gap segments, and dotted areas aspirative segments.
stops. It is seen in Fig. II-B-2 that the initial /g/ of the words gâda, gâta, gâda, and gâta must be placed at the same point relative to the pitch contour. The same holds for the initial /k/ of words kâda, kâta, kâda, and kâta. With this alignment the total k-segment (silent interval+aspiration segment) starts about 40 msec earlier and ends about 40 msec later than the /g/-segment. The release of /k/ occurs at a time corresponding to the middle of the /g/-closure. A similar relation holds between the syllable final /t:/ - /d:/ and /t/ - /d/ pairs.

(Compare gâda and kâda with gâta and kâta, and also gâda and kâda with gâta and kâta.) The voiceless cognate starts 40 msec earlier and ends 40 msec later than the voiced one. In these cases there is no aspirative segment in the /t(ː)/’s, however. (The same relations obtain when the words are aligned not with respect to the pitch contour but with respect to the visible formant transition patterns in the vowels before, during, and after the test words. The data showing this will be presented in a later QPSR.)

It is well known that the first formant transition is less extensive in vowels preceding and following voiceless stops than in those surrounding the corresponding voiced stops. Apparently, the vocal cords stop vibrating before the articulators reach the state of complete closure and start vibrating again only some time after the release. One might therefore hypothesize that the articulatory gesture is identical in the syllable final voiced and voiceless stops.

The articulatory gesture must, however, be different in the syllable initial voiced and voiceless stops (/g/ and /k/) since the aspirative release of the /k/ starts at a time which corresponds to the middle of the /g/-closure. The temporal relations are such, however, that the aspirative element appears to be inserted at the end of an otherwise intact unaspirated voiceless stop gesture. This question needs further investigation.

The relative timing between /v/ and /f/ resembles that between /g/ and /k/. The situation is not equally clear in the case of the /j/- /g/ pair. The latter may be due to the difficulty of segmenting the smooth formant transitions of the
voiced fricatives. Note, incidentally, the displacement of the /d/-segments of words containing fricatives relative to those containing only stops.

Another point of interest is the timing of the vowel segments relative to the pitch contour. The long /a:/ of /gã:da/, for instance, captures the entire falling phase of the pitch contour and also some of the subsequent rise. The short initial /æ/ of /kæt:2/, however, hits only a small segment of the initial portion of the falling phase. In fact, as Fig. II-B-2 shows, if the tonal accents were studied only within the temporal domain of the vowel then a very large number of quite different "allo-pitch contours" would be obtained, at least in the case of the falling tone.

If the different starting frequencies of $f_0$ after initial voiced and voiceless stops and fricatives can be explained as automatic consequences of articulation, then it should be possible to derive a single invariant vocal cord tension contour which would represent the falling tone in bisyllabic non-compound words. The constancy of this curve as evidenced by Fig. II-B-2 suggests that the timing of the articulatory gestures are arranged with respect to the time scale set by the phonatory gesture according to certain rules. Apparently, these rules allow some of the articulatory segments to float on the tonal contour in certain ways. These and the preceding observations lead to a number of questions with which we conclude the present report.

(1) Is there an articulatory time unit, larger than that of the phoneme gesture, the boundaries of which are locked to fixed points on the tonal gesture?

(2) Do the derived-pitch contours of Fig. II-B-2 remain unchanged under phonetically more complex words, i.e. words containing consonant clusters?

(3) Are the pitch contours of Fig. II-B-2 different in compound words?

(4) How does the sentence intonation affect the tonal gesture?

Reference: