Word and sentence intonation: A quantitative model

Öhman, S.

journal: STL-QPSR
volume: 8
number: 2-3
year: 1967
pages: 020-054

http://www.speech.kth.se/qpsr
B. WORD AND SENTENCE INTONATION: A QUANTITATIVE MODEL

S. Öhman

In previous papers \(^{(1)}\) we have suggested a functional model of pitch control in speech production \(^{(2)}\). The present report summarizes the results of some further attempts to explore as well as to constrain the model by comparing it with empirical data. We shall concentrate here on the patterns of control that must be postulated to account for the Scandinavian word tones and that present certain neuro-motor implications \(^{(3)}\). We will then briefly consider the model in relation to tone languages in general.

The discussions that we shall offer below are not meant as a definitive treatise of word and sentence intonation. The hypotheses to be proposed will probably have to be modified in the light of data that have not been available to the author and/or that are yet to be discovered. It is hoped that it will be possible to collect some of these data through experimental use of the tentative model outlined below. The purpose of this paper is thus to stimulate discussion and research.

Model

The main features of the model \(^{(4)}\) are summarized in Fig. II-B-1. The fundamental frequency signal \(f_o(t)\) is synthesized by a mechanism (labeled "larynx model") which accepts three types of input: 1) the time varying "vocal cord tension" \(g(t)\), which is the sum of two components, \(g_s(t)\) and \(g_w(t)\), where \(g_s(t)\) represents the sentence intonation and \(g_w(t)\) the word intonation, 2) an acoustic interaction signal arising from the secondary effects on \(f_o\) caused by fluctuations in sub- and supraglottal pressure \(^{(5)}\), and 3) an articulatory interaction signal deriving from the non-phonatory movements of the hyo-thyroid lever system. These movements are due to certain articulatory gestures of the tongue \(^{(6)}\).

The two signals \(g_s(t)\) and \(g_w(t)\) are the outputs of two filters that may have different properties. These filters are assumed to represent the dynamic characteristics of the mechanical and peripheral-neural components of the laryngeal system.

The inputs to the sentence and the word intonation filters are supposed to be built up from step functions with different amplitudes and onset times. These steps represent the discrete higher-level neural
FUNCTIONAL MODEL OF LARYNX CONTROL

Fig. II-B-1.
events that correspond to the linguistic intonation elements (levels, tones, etc.). In particular, it is assumed that only a finite "library" of step amplitude and timing configurations may be drawn upon in the construction of the \( f_o \)-contour of an utterance. Empirical investigation involving systematic matching of the model to data must decide what these configurations are like.

It is necessary to formulate exactly the properties that should be incorporated into the boxes labeled "larynx model", "word intonation filter", and "sentence intonation filter" in Fig. II-B-1. The particular choices that have been made for the purposes of the present study are indicated in Fig. II-B-2. We will not discuss in detail the physiological facts that motivate these choices here(7). The following points should be mentioned however.

1) The relationship \( f_o = e^G \) is suggested by the fact that the \( f_o \)-contours of the same utterance spoken at different over-all pitch levels (pitch registers) give essentially parallel curves when plotted on a logarithmic frequency scale.

2) The effects of the articulatory and acoustic interactions are disregarded since we will avoid utterances containing voiceless and/or strongly obstructed sounds in the important parts of the intonation contours (8).

3) The form of the impulse responses of the word and intonation filters is assumed to be \( t^n e^{-\gamma t} \) in analogy with the general shape of the tension developed by striated muscle in a single twitch (9). It is to be expected that the mechanical interconnections in the larynx and the characteristics of the neuro-motor circuits do not cause the form of the response function to differ appreciably from that of muscle. This hypothesis is suggested by experience from movement analysis of lip activity in speech (10).

We repeat that the word and sentence intonation filters, which are jointly represented by \( g'(n, \gamma; t) \) in Fig. II-B-2, may have different impulse responses. This is to say that the constants \( \gamma \) and \( n \) must be determined separately for the sentence and word components through comparison with actual data (11). The bottom part of Fig. II-B-2 shows impulse responses for \( n=1 \) and \( n=2 \).

Computer program

The measurements and calculations of the present study have been made by means of a Control Data 1700 as well as a PDP-7 (12) digital computer. The program used for measuring the \( f_o \)-contours of human utterances has been described in STL-QPSR 1/1966 (13). A simplified version has also been written by Tjernlund for the CDC-1700. The numerical simulation of the model of Fig. II-B-2 was implemented by Öhman on the PDP-7 for visual analysis purposes. Another program described in the present issue of QPSR (Sec. IV.B) has been written by Liljencrants (14).
FUNCTIONAL MODEL OF LARYNGEAL CONTROL IN INTONATION

NEURO-MOTOR COMMAND

VOCAL CORD "TENSION"

FUNDAMENTAL FREQUENCY

\[ g(t) = \int_{0}^{\gamma} u_{j}(t-t_{j}) g'(n, x; t) \, dx \]

\[ f_{0}(t) \]

\[ g'(1, \gamma; t) = \gamma^{2} t e^{\gamma t} \]

\[ g'(2, \gamma; t) = (\gamma^{2}) t^{2} e^{\gamma t} \]

Fig. II-B-2. Detailed specification of one of the channels shown in Fig. II-B-1.
In the computer, the measured pitch contour, $f_o(t)$, is compared with the calculated function $F_o e^{g(t)}$, where $g(t) = g_s(t) + g_w(t)$, $F_o$ does not vary with time, and

$$g_x(t) = \sum_i C_i u_{-1}(t-t_i)(1 - (1+\gamma_i(t-t_i))^{-\gamma_i(t-t_i)})$$

$x = s$ or $w$  \hspace{1cm} (Eq. 1)

(The symbols $C_i$ and $\gamma$ will be replaced by $A_i$ and $\alpha$ in connection with $g_s(t)$ and by $B_i$ and $\beta_i$ in connection with $g_w(t)$.)

The criterion of best fit between model and data is based on the mean square percentage error $\varepsilon$, defined by

$$\varepsilon = \frac{1}{T} \int_0^T (1 - \frac{F_o e^{g(t)}}{f_o(t)})^2 dt$$

(Eq. 2)

Programs are available (15) which will automatically adjust the parameters so as to minimize this error when a visually determined gross approximation is specified along with the measured data.

The Scandinavian tonal accents

Computer-aided purely visual matching of calculated functions with measured $f_o$-contours has indicated that good approximations may be obtained even with a small number of input steps. However, the visual criterion does not in general uniquely determine the values of the parameters and the minimum number of necessary steps. This fact indicates that constraints of a linguistic nature must be imposed on the model.

The Scandinavian word tones provide a corpus of data that may be used for this purpose. In the majority of the Scandinavian dialects stressed words are characterized by one of two tonal accents, usually called acute and grave, or Accent I and Accent II, respectively. The linguistic function of these accents is similar in the various dialects in that they usually reflect the phonological structure of the morphemes entering into the word. In Stockholm, for instance, a word has the acute accent if the stem plus the closest suffix contain exactly one vowel. Otherwise it has the grave accent (16). It is, moreover, believed that the Scandinavian accents have a common historical origin (17).

The phonetic manifestations of the accents vary a great deal from dialect to dialect, however. In Standard Danish, for instance, the
acute accent is realized by a rising pitch during, and a glottal stop at the end of, the stressed vowel, and the grave accent consists of a pitch contour which is level during most of the first syllable and then rises abruptly at the end of the syllable (18). In Stockholm, on the other hand, the pitch is rising in acute accented vowels and falling in grave accented vowels, while the situation is almost reversed in Southern Swedish Skåne (19). In many of the dialects of Dalarna (Dalecarlia), finally, the pitch pattern is rising during the first vowel of both the acute and the grave accents, the difference being that the pitch drop starts earlier in the acute than in the grave accented words (cf., for instance, dialects No. 23 through No. 29 of Fig. II-B-3).

Fig. II-B-3 gives a qualitative summary of the pitch patterns of Scandinavian one-word utterances, as established by Meyer (20).

In view of the phonological unity of the accent phenomenon it may be suspected that there is some common physiological basis behind its diverse phonetic realizations. If such a basis can be found, a study of its properties may suggest universal constraints on the intonation model.

Stockholm sentence intonation

The speech material of the present study consists mainly of non-compound Swedish accented words embedded in one of the two frames [seja _____ igen], "to say _____ again" or [de va _____ ja sa:], "it was ____ that I said". In an impressionistic description of the sentence intonation of these frames as spoken in the Stockholm dialect the pitch would be medium high on the first two syllables ([seja] or [de va]), high from the third syllable on, and low on the last and perhaps also on the second last syllable. Using a subjective scale with five levels (0-4) and assigning one level to each syllable of the frame we would write [334 ... 40], and sometimes [334 ... 20] or even [334 ... 00] depending on the particular mode of pronunciation of the speaker (21).

These subjective data indicate that the sentence intonation part of the model should be adjusted in the following manner. The constant $F_0$ is chosen so as to match the pitch of the first two syllables (subjective level 3). Somewhere in the neighborhood of the first syllable of the accented word a positive step is introduced into the sentence intonation filter. A second step of negative amplitude, representing the end-contour, is then added in the neighborhood of the penultimate syllable.
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.
of the utterance (22). To go beyond this point, however, it is necessary to introduce the following TIMING ASSUMPTION: "The positive sentence intonation step starts at the \textit{beginning} of the first syllable of the accented word (23)."

Model commands for Stockholm word intonation

The top part of Fig. II-B-4 shows a typical acute accent pattern (Stockholm dialect) corresponding to the utterance \[\text{de va m\ddot{o}:nen ja s\ddot{a}:}\] (24). Here the beginning of the first syllable of the accented word \[\text{m\ddot{o}:nen}\] coincides with the beginning of the closure for \[\text{m}\] as seen from the acoustic record. The superimposed smooth line in the top left part of the figure (curve marked M) shows the result of introducing two sentence intonation steps (one positive and one negative) in accordance with the considerations just discussed. The composite step configuration (marked I) is shown below the \(f_o\)-contours.

It will be seen from the figure that the model contour fits the measured contour rather badly in the segment immediately to the right of the positive step (25). The error curve (marked E) which represents what is left in the measured contour after the model contour has been subtracted from it, displays a negative dip near the beginning of the first syllable of \[\text{m\ddot{o}:nen}\]. This dip cannot be removed without adding further steps into the model curve. It is a typical feature of the Stockholm acute accent pattern as indicated by an analysis of about a hundred contours, and it is often much more pronounced than the example of Fig. II-B-4 might suggest (26).

The situation is quite similar as regards the grave accent contours. This fact is demonstrated in the lower left part of Fig. II-B-4. Here a positive and a negative sentence intonation step have been introduced in the same manner as was done with the acute accent pattern. The error curve again displays a negative dip which occurs somewhat later in this case than in the previous case. Obviously, the mismatch between model and data cannot be removed without entering further steps into the model contour.

It was noted above that the impressionistic sentence intonation for both \[\text{de va m\ddot{o}:nen ja s\ddot{a}:}\] and \[\text{de va m\ddot{o}:nen ja s\ddot{a}:}\] would be recorded as \(334440\) (or \(3344\times0\) \(0 \leq x \leq 4\)). Since the sentence intonation appears to be level in the accented syllables it may therefore be proposed
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 $\alpha$ and $\beta$ both set to 1000.
that the two residual error contours of Fig. II-B-4 discussed above should be ascribed to the word intonation commands for the acute and the grave accents. The right part of Fig. II-B-4 shows the result obtained by entering a negative pulse (marked IW) of appropriate duration, amplitude and onset time into the word intonation filter simultaneously with the step configuration for the sentence intonation arrived at in the previous analysis (curve marked IS). The error has now decreased to 3 or 4 percent.

Pulse theory of Scandinavian accents

The idea strongly suggests itself that the Stockholm tonal accents should be represented by a suitably tailored negative pulse fed to the word intonation filter at a certain moment relative to the sentence intonation step. This moment is early for the acute accent and late for the grave one (27). This scheme is summarized in Fig. II-B-5 the symbols of which should be self-explanatory.

The model response to inputs of the form of Fig. II-B-5 under systematic variation of the relevant parameters is shown in Fig. II-B-6. The labels indicate the parameters that were varied in each case shown. In the curves labelled A and α the value of B was zero (see Fig. II-B-5 and p. 22 for meanings of parameter names). In the curves labeled B, β, and D the value of A was zero. The curves labeled t_2, finally, represent combinations of a typical sentence intonation step with a typical word intonation pulse, the relative timing of these two events being changed systematically. Fig. II-B-6 should give an idea of the possibilities of the scheme summarized in Fig. II-B-5.

A qualitative comparison of the patterns that can be synthesized using the scheme of Fig. II-B-5 with descriptions of the accentual pitch contours of a wide variety of Scandinavian dialects (cf. Fig. II-B-3) suggests that it may be possible to put all these dialectal manifestations on the same formula.

Model commands for Malmö word intonation

Fig. II-B-7 gives an example from Southern Swedish (Malmö) arranged in the same manner as that of Fig. II-B-4. The utterances of this example are [sɛja môː:n̩ ijen] and [sɛja móː:n̩ ijen]. Note that the end-contour of the sentence intonation differs from that of the
INPUT COMMANDS FOR SWEDISH ACCENTS

\[ \text{C} \mid \text{V} \mid \text{C} \]

SENTENCE INTONATION STEP

\[ t_1 \quad t_2 \]

WORD INTONATION PULSE

---

Fig. II-B-5.
Fig. II-B-6. Model responses to input commands of Fig. II-B-5. Explanation in text.
Comparison of acute and grave accent patterns of the Malmö dialect with model generated curves (cf. text to Fig. II-B-4). The IS step starting at [i] is due to the fact that the latter word was stressed. The glottal stop at the beginning of [i], marked by the letter G both in the data curves and in the IW curves, is discussed in the text (p. 25).
Stockholm speaker. Also, a glottal stop occurs at the beginning of [ijën] in both utterances. The latter seems to be a general feature of Swedish phonology: a syllable beginning in a vowel may be preceded by a glottal stop (28).

Evidently, the word intonation pulses for the Southern acute and grave accents differ from those of the Stockholm speaker in that the grave pulse starts earlier than the acute one. I.e., the temporal order of the Malmö acute and grave pulses is the reverse of that of the corresponding Stockholm pulses. Moreover, the Malmö acute has longer duration, D, and a steeper time course (greater value for $\beta$) than the Stockholm acute does.

These differences probably account for the fact that the subjective impression of the directions of the accental pitch movements are opposite in the two dialects (cf. Fig. II-B-8). In the Stockholm acute the pitch drop due to the negative word intonation pulse is completed during the initial consonant of the stressed syllable and the return of the pitch up to the sentence intonation level occurs during the vowel. This is heard as "rising pitch".

In the Malmö acute, on the other hand, the word intonation pulse starts only after the initial consonant. Thus the pitch will begin with an upward movement during the initial consonant and the very first portion of the vowel in response to the sentence intonation step. The negative word intonation pulse then makes the pitch turn in the downward direction during the later part of the vowel. Apparently, this configuration is perceived as "falling pitch".

Conversely, the onset of the Stockholm grave pulse is late enough to permit the pitch to rise in response to the sentence intonation step during the initial consonant and then to drop throughout most of the vowel so that the impression of a "falling pitch" results. At the end of the first vowel the pitch then turns back up towards the sentence intonation level. The relatively high pitch on the second syllable is correlated with an impression of "tertiary stress".

The Malmö grave, on the other hand, holds back the pitch rise due to the positive intonation step during the initial consonant and part of the initial vowel segment so that a "rising pitch" is heard. In fact, the Malmö grave accent of many speakers sounds very similar to the acute accent of many Stockholm speakers. These relationships are illustrated schematically in Fig. II-B-8.
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.
Perception of tones

It appears from this discussion that the perception of pitch movement in the Swedish accents may be based on a subjective measurement of fundamental frequency during the vowel segments only\(^{(29)}\). This assumption appears reasonable also in view of the possibility of having accented words with all consonants voiceless. In a word such as [stik:at], about 700 msec long, the two voiced segments [i] and [a] may be about 80-150 msec each or even less. Yet the accent is immediately perceivable!\(^{(30)}\)

Model commands for the Danish word intonation

The glottal stop typical of the Standard Danish counterpart of the Stockholm acute accent can be simulated in our model by introducing a fast and brief word intonation pulse late in the stressed syllable\(^{(31)}\) (cf. Fig. II-B-9). The Danish "grave accent", on the other hand, describes a level pitch contour during most of the stressed syllable, followed by a rapid rise/fall pattern that starts just before the medial consonant and reaches its peak in the middle of the second vowel. This pattern is well matched by the command configuration of Fig. II-B-5, if the word intonation pulse is put at the beginning of the stressed syllable.

We must, accordingly, conclude that the order of the Danish word intonation pulses, like that observed in the Malmö dialect, is opposite to the order of the Stockholm word intonation pulses.

Model commands for North Swedish word intonation

It is interesting to contrast this situation with that of Northern Swedish (cf. Fig. II-B-10) where the acute accent pulse either is absent or is placed on an unstressed syllable immediately preceding the accented word. Many of these Northern dialects also have so-called circumflex, i.e. monosyllabic words with the grave accent (cf. [bit:] = "bite!" and [bif:] = "to bite"). These forms have arisen through apocopeation of the final vowel of a bisyllabic grave accented word\(^{(32)}\).

That also the grave accent pulse is earlier in the North than in Stockholm follows from a comparison of the relative heights of the two pitch peaks of this accent: in the North the second peak is, in general, higher than the first, but in Stockholm the first peak is usually higher
Fig. II-B-10. Comparison of North Swedish (Mycelgensjö) accents with model generated curves. Cf. Fig. II-B-4.
"ACUTE ACCENT" : COPENHAGEN

"GRAVE ACCENT" : COPENHAGEN

Fig. II-B-9. Comparison of Danish accent patterns with model generated curves.
Cf. Fig. II-B-4.
than the second. Compare, for instance, "Akzent 2" of dialects No. 3 and No. 48 of Fig. II-B.3. We return to this fact below.

Model commands for Orsa word intonation

The two accents of a typical Central Dalarna dialect (Orsa) are shown in Fig. II-B.11. The scheme of Fig. II-B.5 can be fitted to these data in the simplest way by assuming that no word intonation pulses occur and that the whole contrast depends on the timing of the end-contour. The latter is early for the acute accent and late for the grave accent.

It is also possible to let the terminal contour be equally late in both cases and to enter a word intonation pulse into the second syllable of the acute accented word. If the model constants are chosen appropriately, a pitch contour can be obtained in this way that differs only slightly from that shown in Fig. II-B.11. The significance of this ambiguity will be discussed in more detail below.

PHYSIOLOGICAL INTERPRETATION OF MODEL ELEMENTS

We will now discuss from a physiological point of view the model commands that we have used in synthesizing the accent contours. First, the word intonation pulses will be considered and some data will be presented relating to their neuro-motor correlates in the laryngeal muscles during normal phonation as well as to their acoustic correlates during whispered speech. We will also make a brief comment in this connection on the relation between perceptual prominence and pitch movements in the accented syllables.

We will then turn to the problem of interpreting, in physiological terms, the model elements out of which the sentence intonation is built up. This discussion will lead to certain modifications of the theory presented in earlier sections of this paper. In particular, the sentence intonation commands will be divided up into two parts, the basic phrase contour and the phonatory stress pulses. This modification, though fundamental, does not, however, change our previous analysis of the Scandinavian word tones into a (positive) sentence intonation component with a superimposed (negative) word intonation deflection. The modification is rather to be understood as an extension of the theory building on the earlier results.
Fig. II-B-11. Comparison of Dalarna (Orsa) accent patterns with model generated curves.
Cf. Fig. II-B-4.
After having discussed the physiological correlates of the model elements we shall return to the Scandinavian tonal accents in a more systematic way and propose a tentative and necessarily incomplete theory as to their possible origins and development. Finally, a few comments will be made concerning tone languages outside of Scandinavia.

Glottal stop interpretation of word intonation pulses

The phonetic nature of the Danish "acute accent" suggests that all Scandinavian accents should be understood as variously timed glottal stops, only softer than the Danish ones (33). At the moment, this is an open question.

For one thing it is not quite clear what a glottal stop is, physiologically speaking. It is well known, however, that the muscles of the larynx are innervated by two distinct branches of the X'th cranial nerve, the recurrence and the laryngicus cranialis (34). The latter branch ends in the two parts of the crico-thyroid muscle (pars recta and pars obliqua) while the recurrence branch innervates the remaining intrinsic muscles of the larynx. It is believed that the crico-thyroid muscle is the one mainly responsible for pitch control in singing and in speech (34). It is thought to perform this function by causing a rotation of the thyroid cartilage about its joint on the lower posterior part of the cricoid cartilage thereby stretching or relaxing the vocal cords.

The valve function of the larynx, on the other hand, evident in glottal stops and in the unvoicing gesture of truly voiceless consonants (this unvoicing gesture is apparently a sort of negative glottal stop!) - is probably performed by the various adducing and abducing muscles. There is some evidence of reciprocal inhibition between the crico-thyroid system and parts of the adducing/abducing system, fn (34).

In a small pilot experiment (35) the motor unit activity of the vocalis and the crico-thyroid muscles was recorded by means of thin concentric needle electrodes. The results will be summarized in a later section of this CPSR (sec. II.C). The main finding, however, is a brief phase of inhibition of crico-thyroid activity at the moment where the negative word intonation pulses would occur for the two accents (early for the acute accent and late for the grave). The "ballistic" character of this inhibitory phase is consistent with the glottal-stop theory of the accents.
In view of the relative sizes of the anatomical structures affected by the two laryngeal control systems, one would expect the pitch function to be more sluggish than the valve function. It is therefore interesting to observe that in matching the accent pitch patterns it has invariably been necessary to assign a value of about 8 to the model constant $\alpha$ and values exceeding 12 to $\beta$. Thus, in terms of these parameters, the word intonation filter of the model has to be at least 50% faster than the sentence intonation filter.

A relevant point in this connection is the pronunciation of the words meaning "yes" and "no" in the Stockholm dialect. Normally, these words are pronounced $[\text{jó:}]$ and $[\text{ně:}]$ with the acute accent. But they can also be spoken bisyllabically and with the grave accent, $[\text{jâ:u}]$ and $[\text{ně:e}]$. In emphatic speech they are produced with glottal stops, $[\text{já:}]$ and $[\text{ně:}]$.

**Whispered accents**

Surprisingly enough, the word accent distinction does not disappear in whispered speech. Under this condition the resonances of the vocal tract are excited by a turbulent noise source generated by the incompletely closed glottis. At moderate Reynolds’ numbers the spectral peaks of this noise are tuned to the vocal tract formants. It is well known that the resonance frequencies of a tube open at both ends (both at the lips and at the glottis, in the case of the vocal tract) will be higher than those of a tube closed at one end, provided that the shape of the tube is unchanged. If the word accents involve a brief narrowing of the glottis orifice, a downward movement of the "noise pitch" would therefore be expected. Evidently, a decrease in noise intensity also accompanies the accent gesture. Effects of this type may convey the subjective word accent impression in whispered speech.

**Prominence of accented syllables**

A brief comment is due also regarding the impressionistic assignment of prominence levels to the syllables affected by the word accents. Using a subjective scale of five levels, (0-4), traditional phoneticians put [40] for acute accented words of the Stockholm dialect such as $[\text{mó:něn}]$ and [32] for grave accented words such as $[\text{mō:nen}]$. I.e., the second syllable of the grave accented word is perceived as being more prominent than the corresponding syllable of the acute accented word.
In terms of the pulse theory (refer to upper part of Fig. II-B-8) the acute accent pulse occurs early in the first syllable while the grave accent pulse is late. The laryngeal effect needed to return the pitch to the sentence intonation level after the negative accent pulse has occurred will therefore be greater in the grave accent case, since the pitch contour without the accent pulse would be close to its asymptote near the end of the first vowel. Moreover, this effort must be expended during most of the second syllable in the grave case in contrast to the acute case where it occurs during the first syllable. It may be that the perceived degrees of prominence are at least partly based on these relationships (39).

PHYSIOLOGICAL INTERPRETATION OF SENTENCE INTONATION COMMANDS

We have as yet only studied the tonal accents in sentence positions characterized by essentially constant subglottal pressure, $P_s$, and constant "average pitch level". However, a sentence intonation contour may be subdivided into consecutive phrases each of which is grossly characterized by an initial rise, a flat peak medially, and (usually) a fast drop at the end of both the $P_s$ and the $f_o$-contours (40). (The simple declarative sentence should be regarded as a single phrase.)

If one assumes, as we have done, that $F_o$ is constant (cf. p. 22 et seq) then one is forced to conclude that while the sentence intonation steps are positive in the pre-peak portion of the phrase, they should be negative when occurring after the point where the peak is desired in order to bring about the typical pitch drop at the end of the over-all $f_o$-contour. If this were so, the tonal accents could not be generated by means of a single negative word intonation pulse in the falling part of the $f_o$-contour, for the accentual pitch patterns do not lose their familiar shape in these positions. In particular, the Stockholm grave accent always starts with a rise or at least a strong reduction in the rate of fall in pitch. But no combination of a negative sentence intonation step with a negative word intonation pulse could bring this effect about.

This problem can apparently be solved by assuming that

1) the model parameter $F_o$ is not constant throughout the phrase but describes a slow rising-falling movement typical of the breath-group henceforth to be called the basic phrase contour the detailed properties of which are yet to be determined by empirical measurements.
2) the sentence intonation commands that are added to the basic phrase contour are not steps, as assumed previously, but pulses. These pulses, which we will call phonatory stress pulses, have a duration of the order of the syllable and are always of positive amplitude.

3) a phonatory stress pulse of some non-zero amplitude is centered at the beginning of every stressed syllable according to a strategy that will be described below.

4) the word intonation is simulated by means of negative word intonation pulses in the same manner as proposed previously (p. 25).

Note that requirement 2) guarantees that the appropriate word tone patterns can be synthesized with the model also in the falling part of the over-all intonation contour. Since the phonatory stress pulse always has positive amplitude it will always bring about an upward inflection of the $f_0$-contour. On the other hand, since it is a pulse of finite duration it will only cause a local perturbation on the basic phrase contour so that the over-all form of the latter is not too drastically changed.

Properties of the basic phrase contour

As was noted above, the detailed properties of the deepest of the intonation processes, i.e. the basic phrase contour, must be established by means of phonetic measurements. Given sufficient information about the general properties of phonatory stress pulses and word intonation pulses it should be possible to take $f_0$-contours from human utterances, subtract off the pitch modulations due to stress and to word tones, and thereby obtain the basic phrase contour as a residue (with a superimposed ripple due to noise as well as to articulatory and acoustic interactions). This procedure is illustrated in Fig. II-B-12.

Although the study of the basic phrase contour, as defined above, has not yet proceeded very far, it is nevertheless possible to distinguish three typical patterns; that of the terminative mode, that of the continuative mode, and that of the elicitative mode of pronunciation. The terminative mode is characterized by a steeply falling end-contour that starts close to the end of the latest stressed word of the utterance and is normally used in simple declarative sentences. The continuative mode has a level end-contour and is normally used to indicate that more phrases will follow. The elicitative mode, finally, has a rising end-contour (often followed by a fast drop) and is used in questions and to express surprise.
Fig. II-B-12. Curve I: Measured pitch contour of the utterance "and that morning the man was murdered", spoken in the elicitative mode (Stockholm dialect).

Curve II: Model response to the four phonatory stress pulses shown in curve V. For all of these input pulses the amplitude was .5, the duration .3 sec, and the filter constant $\alpha = 8$.

Curve III shows the result of adding the word intonation pulses of curve VI to the input. The latter pulses are all identical and $B = .8, D = .15$ sec, and $\beta = 15$. Curve IV is the difference between Curve I and Curve III and probably represents the basic phrase contour.

Note the typical elicitative end-contour.
Physiological intensity

The basic phrase contour which is characteristic of the phonological phrase, may perhaps ultimately be related to the breathing cycle which, in speech, is somewhat distorted by the talker's attempt to reduce the rate of change of subglottal pressure so as to maintain a locally constant pressure source for phonation.

But the slow subglottal pressure inflection and the basic phrase contour are not the only phonetic correlates of the phonological phrase. It is known, for instance, that equally stressed syllables are longer in the initial and especially in the final positions of the phrase than they are medially. With this lengthening there seems to follow a general relaxation and neutralization of the articulatory gestures of the syllable. This articulatory relaxation which increases with increasing syllable length is evidently due to a different sort of process than that which is correlated with decreasing stress. We return to this problem below.

Perhaps all these manifestations of the phonological phrase should be viewed as the phonetic correlates of a single physiological process that, for want of better name, might be called physiological intensity, \( I_p \). The speed of response of the whole speech system to phonetic motor-commands would thus be monotonically related to \( I_p \). In particular, the time course of \( F_o \) and \( P_o \) would reflect this physiological intensity rather directly.

The notion of \( F_o(t) \) as an image of a hypothetical physiological intensity function, \( I_p(t) \), suggests a new interpretation of the stress pulses of our model. These pulses would reflect an instantaneous addition of a quantum of physiological energy to the speech production system as a whole. This energy is distributed spatially, over the articulatory, phonatory, and pulmonary channels, as well as temporally, over the time segment of the utterance immediately following the onset of the pulse. This theory could apparently explain a number of superficially unrelated phonetic correlates of stress.

Firstly, the physiological intensity increase in the pulmonary channel will bring about a ballistic contraction of the intercostal muscles thereby producing a brief increase in subglottal pressure, i.e. a pulmonary stress pulse with a consequent increase in acoustic source intensity.
Secondly, the increase of physiological intensity in the articulatory channel would bring about faster and more vigorous movements of the tongue, lips, velum, and jaw, as well as of the articulatory components of the larynx. As a consequence of this, the stressed syllables would be characterized by a lower degree of coarticulatory overlap between the abutting consonant and vowel gestures. Phonetic distinctions would therefore become sharper in these syllables.

Finally, the physiological intensity increase in the phonatory channel will—as already noted—bring about a positive inflection in the f_0-contour through the introduction of a stress pulse at the input of the sentence intonation filter.

The pitch rise due to the phonatory stress pulse is positively correlated with the length of the syllable on which it occurs. However, as was noted above, the duration of equally stressed syllables is smaller in the middle of the phrase, i.e., where I_φ(t) is assumed to have greater values, than at the end of the phrase, where I_φ(t) has smaller values. It is therefore surprising at first, that an increase in I_φ(t), due to the introduction of a stress pulse, should increase rather than decrease the duration of the stressed syllable.

This apparent contradiction may be explained in terms of auditory prominence, a perceptual quality of syllables. It may be assumed that, 

5) the faster the pitch rises in a syllable (or the less steeply it drops), the greater will the perceptual prominence be, and
6) a longer syllable is perceived as more prominent than a shorter syllable.

The lengthening of stressed syllables may now be seen as having two functions, namely, to increase the perceptual prominence of the stressed syllable, and to avoid that the following syllable becomes more prominent than it should be. The second syllable would become too prominent if the pitch rise due to the phonatory stress pulse was not given time to become completed during the stressed syllable but continued to rise in the following syllable.

Physiological energy

The integral of the physiological intensity function I_φ(t) taken over the time interval of the entire phrase might be given the name total physiological energy, E_T. The inverse relationship between syllable
duration (disregarding the effects of stress) and the basic phrase contour, commented on earlier, suggests that, for unstressed syllables, the integral of $I(t)$ over the time segment of the syllable is approximately constant.

Moreover, preliminary experiments indicate that when one of the syllables of an utterance is given emphatic stress two things happen: the over-all pitch level of the remaining syllables drops in proportion to the degree of emphasis, and the pitch modulation due to stress and word intonation in the non-emphatic syllables becomes de-emphasized. This suggests that the physiological energy that goes into the emphasis is subtracted from the energy that otherwise would go into the non-emphatic syllables.

A conservation principle of a similar kind is apparently implied by one of the basic devices in Chomsky's and Halle's generative formulation of American English phonology (48). According to this theory the simplest way of describing the stress patterns possible in the American English phonological phrase is to postulate a (probably universal) type of rule that (a) changes the stress of a syllable to primary stress and at the same time (b) reduces each of the stresses of the remaining syllables of the phrase by one degree. The order in which the primary stresses are introduced into the phrase thus determines the stress pattern as a whole, and this order is a unique function of the morphemic and syntactic structure of the phrase and of the rules. Thus the difference in stress pattern on compound Nouns as compared with Noun Phrases consisting of an Adjective followed by a Noun as well as a great many other prosodic phenomena may be described in a very simple and elegant fashion (49).

On the basis of these hints, the following hypothesis may be proposed.

7) In normal conversation the total physiological energy $E_T$ of a phonological phrase, including the energies of stressed syllables, depends mainly (50) on the number of syllables of the phrase.

**Synthesis strategy for intonation contours**

We are presently experimenting with strategies for the synthesis of $f_0$-contours that obey assumptions (1-7) of pp. 31, 32, 34, and 35. One of these strategies, summarized below, is a phonetic analogue of the Chomsky-Halle sequential procedure mentioned earlier.
We have, so far, only been working with simple declarative sentences in the terminative mode. It is assumed that the basic phrase contour starts with a slow rise from about 90 Hz toward 110 Hz. This rise is turned into a slow drop at a point corresponding to the boundary between the last syllable of the noun-phrase constituent and the first syllable of the verb-phrase constituent of the sentence. This point will henceforth be called the turn-over point. The end-contour, finally, is assumed to start at the beginning of the syllable that immediately follows the latest stressed syllable of the sentence. If the last syllable of the sentence is stressed the end-contour starts in its middle.

The main steps of the procedure are as follows.

SI 1) Assume initially that all syllables are unstressed, calculate a basic phrase contour $F(t)$ for this condition, and calculate the locations of the syllable boundaries on the contour.

This step is carried out using the two assumptions that (with
\[ f_o(t) = e^{g(t)}, \quad g(t) = g_b(t) + g_s(t) + g_w(t), \quad \text{i.e.,} \quad F_o(t) = e^{gb(t)} \]
$g(t)$ is proportional to the physiological intensity $I(t)$, and that, initially, each syllable is given the same physiological energy $E_s$. If the noun-phrase constituent contains $n_1$ syllables and if the latest stress of the sentence occurs on the $a^{th}$ syllable, then the turn-over point, $t_1$, and the time coordinate, $t_2$, of the onset of the end-contour can be found by solving the equations

\[ n_1E_s = \int_{0}^{t_1} g_b(t)dt \]
\[ n_2E_s = \int_{0}^{t_2} g_b(t)dt \]

for $t_1$ and $t_2$. Here $F_o(t) = e^{gb(t)}$ is the basic phrase contour of step SI 1) synthesized according to the principles already described.

The following steps involve a perturbation of the basic phrase contour obtained from SI 1) - both with respect to frequency values and with respect to the durations of the syllables - according to the distribution of stresses within the phrase. The main condition is that the total energy of the phrase remains unchanged. Thus,

SI 2) Specify the sequence in which stresses are to be introduced into the phrase.

SI 3) Calculate the energies of the stresses according to a procedure that will be described in a moment.
SI 4) Calculate the amplitudes of the phonatory stress pulses using the result of SI 3).

SI 5) Recalculate the basic phrase contour and resegment it into syllables.

SI 6) Introduce phonatory stress and word intonation pulses.

The procedure referred to in step SI 3) is as follows. After step SI 1) each syllable has been given the constant energy \( E_s \). If the phrase contains \( n \) syllables, then the total energy \( E_T \) is equal to \( nE_s \). To obtain the amplitudes of the phonatory stress pulses and the new durations of the syllables, we introduce the concept of stress. A stress is a quantum of energy, \( E_i^k \), associated with a syllable at a certain stage of the calculation. The stresses are introduced one at a time. For \( k \geq i \), \( E_i^k \) refers to the value (at the \( k \)'th stage) of the energy quantum introduced at the \( i \)'th stage. The subscript, zero, however, is reserved for the energy of the basic phrase contour. Thus \( E_i^0 \) is the energy of this contour at the \( k \)'th stage.

The procedure indicated in SI 3) may now be summarized by means of the following formulas (a sample derivation is given on p. 38).

\[
\begin{align*}
(\text{SI 3:1}) & \quad E_o^0 = E_T = nE_s \quad \text{and} \quad E_i^0 = 0 \quad \text{for} \quad i > 0 \\
(\text{SI 3:2}) & \quad E_i^{k+1} = \text{constant for} \quad k > 0 \\
(\text{SI 3:3}) & \quad E_i^{k+1} = c(1-2^{-k})E_i^k \quad \text{for} \quad 0 < i < k+1, \quad 0 < c \leq 1 \\
(\text{SI 3:4}) & \quad E_o^{k+1} = E_T - \sum_{i=1}^{k+1} E_i^{k+1}
\end{align*}
\]

Thus, as stresses are introduced one by one into the phrase, the energies of the basic phrase contour and of the stresses introduced earlier are successively reduced. In particular, the energy \( E_i^{k+1} \) of the \( i \)'th stress after the \( k+1 \)'st stage (\( k+1 > i \)) is only a fraction \( c(1-2^{-k}) \), \( 0 \leq c \leq 1 \), of the value it had after the \( k \)'th stage.

Suppose the calculation is terminated after the \( m \)'th stage. We may now determine for any given syllable of the phrase how many stresses have been associated with it during the calculation as well as the energies of these stresses. The sum of these energies is the energy that goes into the phonatory stress pulse to be introduced on the syllable in question. This is what we need to carry out steps SI 4) and SI 5).
Example of Derivation of Stress Energies

<table>
<thead>
<tr>
<th>SEQUENCE OF SYLLABLES IN THE PHRASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAGE</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>E₃</td>
</tr>
<tr>
<td>E₄</td>
</tr>
</tbody>
</table>

Energy to be added to each stressed syllable:

\[
E₁ = c^3(1-2^{-1})(1-2^{-2})(1-2^{-3})E₀^1 \\
E₂ = c^2(1-2^{-2})(1-2^{-3})E₀^2 \\
E₃ = c(1-2^{-3})E₀^3 \\
E₄ = E₀ = E₀^2 = E₀^1 = \text{const.}
\]

\[
\frac{E₄}{E₄} = 2 \quad c^{-1} = 2.22 \\
\frac{E₃}{E₄} = 2.7 \quad c^{-2} = 3.32 \\
\frac{E₃}{E₄} = 3.05 \quad c^{-3} = 4.18
\]
The energy of the phonatory stress pulse is proportional to its amplitude times its duration. Moreover, it is assumed that a stressed syllable is lengthened in proportion to the energy of the phonatory stress pulse placed on it. These conditions determine both the amplitude and the duration of the stress pulse and the duration of the stressed syllable.

The basic phrase contour is now recalculated. The turn-over point and the onset of the end-contour are moved rightwards so that the stressed syllables can be lengthened by the amounts just established. The basic phrase contour so obtained is then multiplied by a constant chosen so that the integral of the resulting contour over the entire phrase is equal to $E^m_o$. Finally, step SI 6) is carried out.

A few comments should be made regarding the synthesis strategy outlined above.

1) The strategy does not impose any limit on the number of stresses that may be introduced into a phrase. However, the "older" a stress is, the smaller will its energy be. A stress introduced five or six stages ago will be almost completely annihilated. This feature satisfies Chomsky and Halle's requirement that "it is necessary to formulate a principle for interpretation of phonetic representations that nullifies distinctions that go beyond a certain degree of refinement".(51)

2) It is not difficult to modify the process so that a number of primary stresses are introduced, so to speak, in parallel, already at step SI 1). This would make the correspondence to the Chomsky-Halle system closer. Given this modification one may also experiment with the possibility of letting the stress adjustments following the initial step involve a redistribution of the initial stress energies only, leaving the energy of the basic phrase contour invariant.

3) A complete synthesis strategy must probably operate with rules that assign energies to syllables also on the basis of the number of phonemes they contain and the varying effort needed to produce these phonemes.

4) We have considered only the phonatory channel but the articular and pulmonary ones are equally important. It may be, for instance, that languages differ with respect to the weights they give each of these channels in realizing stress. Thus, the pulmonary correlates of stress seem to be weak in Japanese, and in American English the articulatory correlates are apparently strongly emphasized as compared with the phonatory correlates.

5) Finally, if an adequate synthesis strategy can be constructed along the lines sketched above, it will be valuable as a tool in the linguistic study of intonation since it may be possible to work out analysis-by-synthesis procedures by means of which the stress patterns of tape recorded human utterances can be measured in a meaningful way.(52).
The Scandinavian accent orbit

Let us now reconsider the Scandinavian word tones in the light of the modified concept of sentence intonation that was discussed in the previous section. Our model now provides us with three types of building block in terms of which the $f_o$-contours of human utterances are to be analyzed. These building blocks are a) the basic phrase contour, b) the (positive) phonatory stress pulses, and c) the (negative) word intonation pulses. We have also suggested that the basic phrase contour is related to the breathing cycle, that the stress pulses reflect the insertion of a quantum of physiological energy into the speech production system as a whole, and that the word intonation pulses correspond to more or less tense glottal stops.

Since we may assume that the basic phrase contour is level in the neighborhood of the test word of utterances like [də va mə:nən ja sa:], etc., the command configuration of Fig. II-B-5 that was suggested as the basic formula for all Scandinavian word tones, need be changed only in one respect. The "sentence intonation step" of amplitude $+A$ is now replaced by a phonatory stress pulse. This change is equivalent to adding a step of amplitude $-A$ to the sentence intonation command of Fig. II-B-5 near the second consonant (cluster) of the syllable.

In all cases that we have analyzed (Figs. II-B-4, II-B-7, II-B-9, II-B-10, and II-B-11) the off-set of the phonatory stress pulse would be very close to the end-contour of the basic phrase pattern. The negative sentence intonation step that was previously used to generate the terminal contour may therefore be regarded as including both the off-set of the phonatory stress pulse and the end command of the basic phrase contour. The modification of the model just introduced consequently does not alter our earlier conclusion that the Scandinavian word tones may be simulated by means of appropriately timed word intonation pulses of negative amplitude.

The pulse character of the phonatory stress command suggests a very natural interpretation of the difference between the Stockholm and the North Swedish grave accent patterns. As was noted earlier, the second peak of the Stockholm grave accent has a lower frequency value than the first peak, whereas this relationship is reversed in many North Swedish dialects. This difference may be obtained in the model by putting the accent pulse slightly earlier or slightly later than the
pitch peak that would result from the phonatory stress pulse alone. This effect is illustrated in Fig. II-B-13.

We have already noted that the difference between the Orsa acute and grave accent patterns may be simulated by entering a word intonation pulse close to the beginning of the second syllable of the acute accented word while the grave pattern is synthesized by means of the phonatory stress pulse and the basic phrase contour only. It is interesting that the same statement holds for some of the (geographically distant) Gotland dialects. According to most authorities, the Central Dalarna and the Gotland Dialects are closer to Proto-Scandinavian than any of the other modern Scandinavian tongues also as regards phonological structure and vocabulary.

To judge from Meyer's data some of the Dalarna dialects show a different type of contrast between the acute and the grave patterns (cf. e.g. dialect No. 18 of Fig. II-B-3) than those observed in Orsa. Here it is more probable that the acute pattern should be synthesized by means of the phonatory stress pulse and the basic phrase contour only, while the grave pattern would be obtained by also entering a word intonation pulse at the beginning of the first syllable of the accented word.

If we take the Rättvik dialect (No. 27 of Fig. II-B-3) as our starting point we may accordingly visualize two ways of sharpening the accent contrast. One way is to enter a word intonation pulse on the falling tail of the acute contour. We may then gradually move this pulse from the right into the first syllable of the word. The other way is to enter a word intonation pulse on the rising ramp of the Rättvik grave pattern. We may then gradually move this pulse from the left toward the end of the first syllable of the word.

In Fig. II-B-14 we have rearranged a part of Meyer's data so as to demonstrate that dialects may be found that correspond to the successive stages of these two gradual processes. The Rättvik dialect will be found in the leftmost part of the "accent orbit". Following the orbit downward from this point on corresponds to the first mentioned of these two processes, and the second process is represented by the opposite direction. (In each subgraph of Fig. II-B-14 the left curve represents the acute accent and the right curve represents the grave pattern.)
Fig. II-B-13. Model outputs for different combinations of input commands.
Curve I: Phonatory stress pulse + terminative end-contour.
Curve II: Terminative end-contour only. Curve III: Word intonation pulse + terminative end-contour. Curves IV-VIII: Phonatory stress pulse + terminative end-contour + word intonation pulse moved stepwise to the right. Note difference in height of the two pitch peaks of curves V and IV.
Fig. II-B-14 was arranged on purely phonetic principles suggested by our model, and not all dialect pairs that are adjacent in the figure are also contiguous geographically. In fact, the dialects do not appear to be distributed along a single geographic orbit, but along several. It is too early to make definitive statements about the exact geographic course of these orbits since a sufficiently dense sampling of the dialects has not yet been carried out over the whole area.

However, Meyer himself emphasized (54) the dialectal continuity along a line that starts in Central Dalarna och moves South-East into Uppland. Examination of Meyer's data suggests that this line divides itself into a Northern and a Southern branch in Uppland. Branch points are also indicated in other parts of the area.

The phonetic importance of investigating these relationships in detail should be stressed. The hypothesis of an accent orbit together with the possibility that there are chains of geographically contiguous dialects that correspond to various sectors of the accent orbit of Fig. II-B-14 implies, for instance, that the acute and the grave accent pulses do not move independently of each other as the dialects develop, but that certain temporal and other qualitative constraints are obeyed so that dialects not in contacts may develop in similar ways for intrinsic reasons. The determination of these constraints may aid our understanding of speech communication in general.

On the origin of the Scandinavian accents

We must ask now if our intonation model can throw any light on the historical problem as to how the Scandinavian tonal accents may have arisen. Needless to say, the remarks that we shall make on this topic are speculative. Moreover, historical linguists disagree about the date and the exact phonological circumstances of the formation of the accent distinction. Below we shall, however, accept the essentials of the views set forth by Ofstedal (55). The discussion offered here may be viewed as an illustration of the central details of his theory in terms of the model proposed in previous sections of this paper.

At the time of Scandinavian linguistic unity (about 500 B.C.) the pitch of stressed syllables (normally the first syllable of the word) would be characterized simply by a phonatory stress pulse. The language distinguished between two types of syllable, namely, long
(CV:C or CVCC) and short (CVC). Both types could be stressed. In the interior of the phonological phrase the tonal difference between stressed short and stressed long syllables would probably be slight. As illustrated in Fig. II-B-15 (parts I and II) the difference would be more marked in phrase final position, however.

In drawing Fig. II-B-15, it has been assumed that the phonatory stress pulse (marked SP) has a fixed duration and amplitude, and that it starts at the beginning of the first consonant of the stressed syllable. Moreover, it has been assumed that the tail of the basic phrase contour (marked BPC) starts on the first consonant of the syllable immediately following the latest stressed word of the phrase. I.e., the end-contour is tied to the final word boundary of the latest stressed word.

At the stage illustrated by parts I and II of Fig. II-B-15 the second syllable of a bisyllabic word with a short first syllable was probably perceived as more prominent than the second syllable of a word with long first syllable. In the former case the pitch is rising and in the latter case it is falling on the second syllable. It is likely, however, that the speakers felt the difference in duration of the first syllable as being more important, since this difference would dominate in the interior of the phrase.

After this period three developments took place in the following order.

1) **Syncopation.** A syllable following a long stressed syllable inside a word was shortened. In particular, if the second syllable was long it became short, and if it was short its vowel disappeared. Later the same kind of second syllable shortening took place after a stressed short syllable. In phrase final position this caused the end-contour of the basic phrase pattern to move with the word boundary closer to the first syllable of the word.

2) **Word boundary shift.** In noun phrases the enclitic definite article, having been a free morpheme (a postponed definite pronoun), became part of the word, i.e., the word boundary was moved one syllable to the right. Also, syllabic word final consonants or consonant clusters that had formed as a result of the syncopation were amplified by means of a svarabhakti vowel. This also had the effect of delaying the word boundary and the end-contour.

3) **Leveling.** Short stressed syllables were lengthened.
Fig. II-B-15. Hypothetical pitch patterns to illustrate possible origin of Scandinavianaccenta. Explanation in text.
Fig. II-B-16. Hypothetical pitch patterns to be compared with East and South East Asian tones.
Parts III and IV of Fig. II-B-15 illustrate the effect of the syncopa-
tion on bisyllabic words with long and short initial syllable. Because
the word boundary has moved to the left the end-contour will truncate
the pitch rise due to the phonatory stress pulse in monosyllabic words.
In words that remained bisyllabic after the syncopation period, on the
other hand, (these words would still have the pitch patterns of parts I
and II of Fig. II-B-15) the pitch is relatively high on the second syl-
lable. A tonal contrast like that of I versus III or II versus IV of Fig.
II-B-15 obtains, as was noted earlier, in many Central Dalarna dia-
lects of to-day.

Since the tonal contrast of I versus III and II versus IV would al-
ways arise in phrase final position owing to the end-contour rule, the
speakers of the time immediately after the syncopation period probably
came to associate an allophonic pitch distinction with the difference
between monosyllabic and bisyllabic words. When the tendency to in-
corporate the enclitic definite pronoun into the word, and to introduce
a svarabhakti vowel into syllabic final clusters, began to make itself
felt, the tonal distinction in phrase final position tended to become the
only surface sign of the deep structure difference between mono- and
polysyllabic words.

If the end-contour simply followed the word boundary one step to
the right, then the pitch contrast would have disappeared (as it seems
to have done in Icelandic and Faroese). If, on the other hand, the
pitch contrast had established itself sufficiently firmly in the linguistic
consciousness of the speakers, then there may have been a tendency
to emphasize this contrast in one way or another.

Now, it is not natural to use a sentence intonation feature, like
the end-contour, to maintain a distinction between words. Normally,
words differ by having different phonemes. It is therefore logical to
expect that the word boundary shift should have been accompanied by
a shift of the end-contour, and that, at the same time, a phoneme-like
feature, such as a word intonation pulse, should have been introduced
into the word so as to sharpen the tonal distinction that had established
itself previously.

This sharpening can be made in one or both of two ways. If the
abrupt pitch drop on the syllable following monosyllabic words (cf.
part III, Fig. II-B-15) was perceived as the marked feature of the
opposition, then a word intonation pulse may have been entered near the beginning of the second syllable of these words as illustrated in part V of Fig. II-B-15. On the other hand, if the relatively high pitch on the second syllable of polysyllabic words (cf. Part I of Fig. II-B-15) was perceived as the marked feature, then a word intonation pulse may have been inserted at the beginning of the first syllable of these words as illustrated in part VI of Fig. II-B-15.

Different dialects would have chosen different combinations of these possibilities. In later stages of the historical development, perhaps in conjunction with the leveling of short syllables, the word intonation pulses must have started to drift in the various dialects, the acute, "monosyllabic" pulse towards the beginning of the first syllable, and the grave, "polysyllabic" pulse towards the end of the first syllable along the lines of the Scandinavian accent orbit. (Cf. parts VII and VIII of Fig. II-B-15)

It should be noted that our theory also postulates a certain tonal distinction at various stages (before the leveling period) between acute accented words with long and short initial syllable as well as between grave accented words with long and short initial syllable. In certain modern dialects that preserve the syllable length distinction (Sollerön, for instance) the accent on short syllabic words is somewhat different from the ordinary grave and acute patterns. It is possible that this feature is related to the processes described in Fig. II-B-15.

Other tone languages

It may at last be worth while to consider very briefly the possibilities of our intonation model in relation to tone languages other than those of Scandinavia. It is reasonable to expect that the word tones of these languages also are superimposed on an underlying basic phrase contour. Whether or not the tones of all tone languages are best described by means of a simple negative pulse fed to the input of the word intonation filter is, of course, an open question at the moment. Even with this restriction imposed on the model, however, a very great number of sharply different tone patterns can be generated.
Suppose that the following constraints were valid for any given tone language:

1) The phonatory stress pulse can only start at the onset of the syllable.

2) Only two amplitudes are allowed for the phonatory stress pulse, zero and a certain non-zero value, $A$.

The negative intonation pulse can only have the following discrete properties:

3) the amplitude is zero or has a fixed non-zero value, $B$;

4) the duration is either short or long;

5) the model constant $\delta$ is either large or small;

6) the word intonation pulse can only occur near the initial consonant(s), near the middle of the vowel(s), or near the final consonant(s) of the syllable.

A language employing all the possibilities that are open under these constraints would have 26 different tones (when $B = 0$ the distinctions under 4), 5), and 6) above become irrelevant). Six of these are shown in Fig. II-B-16. They seem approximately to fit the description of some of the more common tones of the East and South-East Asian language area (58).

It is interesting in this connection to note certain systematic relationships between the Hakka, Foochow, and Pekingese dialects of modern Chinese. In a group of historically related words, the final consonant is a voiceless stop in Hakka, a glottal stop in Foochow, and zero in Pekingese (59). However, these words have a low falling tone in the Pekingese dialect, and specialists believe that this tone has developed from an earlier glottal stop which in turn developed from a (glottalized) final stop consonant. A process like this which seems to be partly inverse to the development of the Danish glottal stop (stød) from a word tone, is nevertheless entirely consistent with the intonation model proposed in this paper.

SUMMARY

In the present paper a quantitative model of larynx control during speech production has been described. The input commands are configurations of simple step functions fed to the model over two channels, the sentence intonation filter and the word intonation filter.
In order to find further constraints to impose on the model for purposes of empirical adequacy, the Scandinavian grave/acute accent opposition was analyzed by fitting curves generated with the model to empirically measured $f_o$-contours.

It was found that the salient features of these intonation patterns in simple utterances of a number of dialects can be simulated by means of a single positive step as input to the sentence intonation filter and an appropriately timed negative pulse as input to the word intonation filter (cf. Fig. II-B-5). It was proposed, tentatively, that this analysis is valid for all Scandinavian dialects.

We next turned to the question as to how the model elements should be interpreted in physiological terms. Regarding the word intonation channel the hypothesis was proposed that the Scandinavian tonal accents are a sort of laryngeal consonants, not unlike glottal stops, that are coarticulated with the sentence intonation as well as with the "segmental" gestures of stressed syllables. The sentence intonation commands, on the other hand, turned out to be decomposable into a basic phrase contour and a sequence of phonatory stress pulses. It was suggested that these constructs reflect an underlying process termed physiological intensity and that stress should be understood as the addition of a quantum of physiological energy to the speech production system as a whole. This energy is distributed (possibly unevenly) over the pulmonary, phonatory, and articulatory channels. In the phonatory channel the stress energy manifests itself as a phonatory stress pulse at the input of the sentence intonation filter.

In this connection, possible energy conserving principles regarding the phonological phrase as a whole were considered. Also, a synthesis strategy for sentence intonation was discussed. This strategy has certain properties in common with the transformational cycle of Chomsky and Halle's theory of phonology.

Having developed these concepts we returned to the Scandinavian accents. An examination of Meyer's data indicated that a series of dialects may be found that display certain systematic relationships with respect to the relative locations of the hypothetical acute and grave accent command pulses. When adjacent members in this series are compared, the accent pulses appear to be cotranslated a small step either to the left or to the right. This relationship was termed the Scandinavian accent orbit.
Consideration of the accent orbit together with known facts of Scandinavian linguistic history suggested an hypothesis about the origin of the word tone distinction. Briefly, this distinction may have arisen in phrase final position as a result of the successive movements of the terminative end-contour.

Finally, a few remarks were made about the generative power of the model with respect to tones observed in non-Scandinavian languages.

The intonation model summarized in the present paper makes it possible to collect systematic quantitative information on the tonal as well as other prosodic events. Work involving close comparison of the model with empirical data is in progress.

ACKNOWLEDGMENTS

I have profited from valuable discussions with G. Fant, B. Lindblom, J. Lindqvist, E. and L. Gårding, K-H. Dahlstedt, R. Leanderson, G. Malmqvist, and A. Ellegård to all of whom I want to express my sincere thanks. I am alone responsible for the ideas and hypotheses presented in this paper, however. I am also grateful to Mrs. S. Felicetti for her expert editorial assistance.

FOOTNOTES


(2) I am grateful to Dr. Philip Lieberman who in many interesting discussions during the academic year 1963-64 drew my attention to the problems of intonation. Cf. Lieberman, P.: Intonation, Perception, and Language (Cambridge, Mass. 1967).
footnotes


(4) The model described here is a slightly revised version of the one presented in fn (1 b). A second revision will be proposed as a result of discussions to follow later in this paper.


(6) This factor has been unsufficiently studied. Cf. fn (1 b) and Faaborg-Andersen, K. and Sonninen, A.: "The Function of the Extrinsic Laryngeal Muscles at Different Pitch", Acta oto-Laryng. 51 (1960), pp. 89-93.

(7) A fuller treatment will be given in a forthcoming publication.

(8) The measurements of Ladefoged, Öhman and Lindqvist, and Ventsov (cf. fn (5)) indicate that the fundamental frequency may vary by 0.16 Hz/cmH2O if only the pressure-drop across the glottis is changed and "everything else is constant. Lieberman gives a somewhat higher value. According to our own measurements, during stressed syllables of normal conversational speech, the subglottal pressure increases by at most a few cm H2O and the increase in pitch due to this factor would therefore be negligible in comparison with the pitch movements caused by larynx muscle adjustments. The effect of pressure fluctuations may be considerable in voiced obstruents, semi-aspiratives, and during the terminal phase of the sentence intonation, however. The tonal configurations studied in the present paper have been embedded in a sentence frame in a position where the subglottal pressure may be assumed to be essentially constant except for minor fluctuations due to stress. In a more complete treatment a pressure dependent correction factor must be introduced into the model calculations in the form of an "acoustic interaction signal".

footnotes


(11) It appears that the best results are obtained with n ≥ 2 for most speakers. All examples of the present paper have been calculated with n = 1, however.

(12) The PDP-7 computer belongs to the Department of Automatic Control, Royal Institute of Technology (KTH), Stockholm. Prof. L. von Håmos' kind cooperation is gratefully acknowledged.


(14) I wish to express my gratitude to J. Liljencrants whose cooperation in the programming phase of this work simplified my efforts substantially.

(15) The uniqueness question for solutions obtained by means of the automatic iterative procedure will be dealt with in a later publication.


(17) Kock, A.: Språkhistoriska undersökningar om svensk accent, Part I (Lund 1878) and Part II (Lund 1884).

Hesselman, B.: Huvudlinjer i nordisk språkhistoria (Uppsala 1948 - 1953).


(18) The Danish glottal stop has been described from various points of view by

Smith, S.: Stødets i danske rigssprog (Copenhagen 1944).

Martinet, A.: La phonologie du mot en Danois.


(19) Malmberg, B.: op.cit.

Hadding-Koch, K.: op.cit. gives spectrographic illustrations.

(20) Meyer, E.A.: op.cit., Teil II.

(21) This analysis, which is an account of my own impressions, purposely disregards the tonal contours of the word accents.

(22) The end-contour can of course be different in different dialects and in different sentence types. Any such contour can be synthesized by means of an appropriately chosen step configuration fed into the sentence intonation filter. In order to purify the effects of the word intonation we have tried to choose frames with maximally simple sentence intonation, however. It so happens that a negative step introduced at the end of the penultimate syllable of the frame referred to in the text suffices to match the Stockholm data satisfactorily. Questions regarding the end-contours will be discussed in more detail on p. 30 et sqq.
The sentence intonation source of the model is assumed here to generate steps only. As a result of the discussion of the physiological meaning of sentence intonation on p. 30 et sqq., we will later replace this assumption by one stating that the sentence intonation source generates positive pulses only. These pulses will also be assumed to start at the beginning of the stressed syllable.

The acute accent will be denoted by "'", and the grave accent by "".

A sharp and brief dip in the measured \( f_0 \)-contour usually occurs during the [v] of [d\(\text{ev}a...\)] of the frame. This dip is probably caused by the increased intraoral pressure during this consonant (fn (1 b)). During the [s] of [... \(\text{ja s\(\text{\`a}\))}\.]) \( f_0 \) is of course undefined.

Meyer, E. A.: op. cit., states that the pitch drop at the beginning of acute accented syllables of the Stockholm dialect represents an influence from Southern Swedish.

E. Haugen, in personal communication, has suggested to me that the East Norwegian acute accent should be identified with the sentence intonation and that the grave accent is a delayed version of the sentence intonation. For the Swedish dialects that I have had experience with so far, it seems better to postulate that the sentence intonation step is fixed at the beginning of the syllable and that the timing (and other parameters) of the word intonation pulse is responsible for the tonal contrast.


The role of this as well as of other types of juncture has been extensively studied by Gårding, E.: "Internal Juncture in Swedish", (thesis, Lund 1967).

In synthesis experiments B. Malmberg has noted that the impression of a grave accent or an acute accent can be obtained by moving a \( f_0 \)-peak within the span of the first syllable of the accented word.


The word intonation pulse at the end of the second syllable of the two Danish test words represents a clearly audible juncture at the beginning of [... \(\text{ja s\(\text{\`a}\))\.}]\. This juncture is similar to the one observed in the Malmö utterances of Fig. II-B-7. Cf. also fn (28).

footnotes

(33) Swedes imitating Danish tend to exaggerate the glottal stops, however. This may be because, as was noted above, tense glottal stops occur in Swedish before stressed syllables beginning in a vowel. The accents are lax in comparison with these glottal stops.


The conception of larynx function briefly summarized here was explained to me by B. Sonesson. I am grateful to him for many illuminating discussions on this topic.

(35) The experiment was carried out at the Central Neurophysiological Laboratory, Karolinska Sjukhuset, Stockholm, by Drs. A. Mårtensson, R. Leanderson, and A. Persson. I am grateful to them for their willingness to cooperate. A more complete description of the methods and procedures used, will be given by Leanderson in a forthcoming publication.


Segerbäck, B.: "La Réalisation d' une Opposition de Tonèmes dans des Dissyllabes Chuchotés" (Lund 1966).


(38) This notation, which has been adopted by Svenska Akademiens Ordbok, has been discussed by C-C. Elert (see fn (3)).

(39) Elert, C-C.: op. cit., p. 139, states that "the difference in duration between words with (acute) accent I and (grave) accent II is analogous to the differences in intensity and the fundamental pitch in the two types of word. The rapid decrease in over-all intensity and the fall in the fundamental pitch in the stressed syllable of (grave) accent II words are accompanied by a shorter duration of the vowels in that syllable".

It may be added that the final syllable of grave words is somewhat longer than that of acute accented words. It is as if the speaker waits for the pitch to return to the sentence intonation level and therefore prolongs the syllable in which this return occurs (first syllable of acute and second syllable of grave words). The relative lengths of corresponding syllables of acute and grave words may of course also contribute to the perceived differences in prominence as well as to the perceived difference in accent in whispered speech.

It is interesting in this connection to note that North Swedish dialects with monosyllabic grave accented words (circumflex) display a duration relationship that is opposite to that of the Stockholm dialect. I.e., the syllable with circumflex is longer than that with the acute accent. (Dahlstedt, K-H.: op. cit., p. 156), yet the native speakers do not "feel" a phonological length contrast. These facts are consistent with the idea that the syllable is lengthened because the speaker waits for the pitch to return to the sentence intonation level.
The over-all intonation contour of the phonological phrase is discussed in great detail by P. Lieberman, op. cit. Although certain aspects of this underlying contour probably are universal it is not unlikely that the details of its shape could vary from language to language (and even from speaker to speaker) being constant for any given language (or speaker).

These laryngeal commands should not be confused with the pulmonary stress pulses that may be observed in the intercostal muscle activity (cf. fn (44)).


In other words, the two syllables immediately following the onset of the phonatory stress pulse may be given different relative prominence simply by adjusting their durations.

Chomsky, N. and Halle, M.: Sound Pattern of English (forthcoming). Prof. Halle has been kind enough to let me see parts of the manuscript of this book before the publication.

In Swedish, the difference between [vär:m kör:v]NE (Adjective + Noun), and [vär:m + kör:v]N (Compound Noun), where 1 denotes primary stress, and 2 secondary stress, would be due to the circumstance that the last primary stress introduced in the transformational cycle, was put on [kör:v] in the first case, and on [vär:m] in the second case.

Future research will show whether ET depends on the number of syllables only or on the number of syllables plus the number of primary stresses introduced at the beginning of the transformational cycle.


It is in my opinion quite premature to conclude that it is impossible to expect a complete correspondence between the records of modern phonetics and the elements and processes postulated in a systematic linguistic theory. In fact, some of the most recent developments in phonetics indicate that correlations of this sort may be successfully established.

As I see it, it is not only possible but necessary to continue work along these lines. First the introspective skills of the auditory phonetician must be translated into objective physical
measurement techniques. Then it may be possible to disambigu- 
guate and sharpen these skills beyond the limits of subjective 
intuition. In this way we may succeed in establishing a sci- 
entific instrument by means of which phonological theories can 
be put to objective test. Naturally, in the initial stages of this 
work our phonetic experiments must be guided by phonological 
theory. As phonetic theory develops, however, it should be 
increasingly feasible to substitute objective phonetic measure- 
ment for impressionistic methods wherever the latter are in- 
determinate.

(53) The chronology of the Scandinavian languages has been dis- 
cussed by E. Haugen in Language 1949, p. 307.
(55) Ref. in fn (17).
(56) Pike, K.: "Tone Languages", Univ. of Michigan Publications, 
Linguistics 4 (1948).
(57) Chang, N.C.T.: "Tones and Intonation in the Chengtu Dialect", 
Abramson, A.: "The Vowels and Tones of Standard Thai", 
(59) Forrest, R.A.D.: op.cit.
Fig. II-B-14. Selection of accent patterns from material of Fig. II-B-3 to suggest cotranslation of acute and grave word intonation pulses.