Predicting physical aspects of English stress

McAllister, R.

journal: STL-QPSR
volume: 12
number: 1
year: 1971
pages: 020-029

http://www.speech.kth.se/qpsr
C. PREDICTING PHYSICAL ASPECTS OF ENGLISH STRESS*
R. McAllister

1. Problem

This paper is a discussion of an attempt to discover correlation between syntactic structure and stress contours in American English.

The problem was made more tractable within this framework by using the recent and well known work of Chomsky and Halle in "The Sound Pattern of English" (1968) as a point of reference. Right branching structure and the rules of stress pertaining to such structures are the object of inquiry in this study. Specifically, the study is concerned with the operation of the Nuclear Stress Rule on right-branching structures (pp. 17-18 in Chomsky and Halle, 1968) and the consequences of this operation in the speech wave.

Correlation has been sought between the output of the Nuclear Stress Rule when applied to a right-branching structure and the fundamental frequency, intensity, and duration parameters in the speech wave. In other words: what do the stress contours that are the outputs of this rule mean in terms of the phonetic facts? The correlation between the output of these rules and the phonetic facts may have important implications in a larger discussion of the role of the phonetic representation in the phonological component of the grammar. This linguistic application of the results of this study will be mentioned later in this paper, at which time the view of the role of the phonetic representation presented in Chomsky and Halle (1968) will be discussed.

Also included in this paper is an attempt to discover regularities in the aforementioned parameters that might lead to the formulation of rules that more adequately express the relevant factors in the correlation between surface structure and the speech wave.

2. Experimental procedures

2.1 Method

Special attention was given to the construction of the test phrases. The syntactic structure has been strictly defined and methodically expanded. Starting with a simple two-constituent tree 'big pants', a right branching tree was built up by expanding the construction to the left.

---

* Thesis work associated with the Dept. of Phonetics, Stockholm University
With this procedure, a range of stress levels generated by the NSR (Nuclear Stress Rule) was obtained for each word in the syntax. In other words, the syntax was expanded in the same way as the transformational cycle operates on a right-branching structure through the NSR.

Fig. 1-C-1 is a summary of the material used in this study presented in terms of its derivation using the procedure described above. Sentence A represents the syntax in its longest form and below sentence A is the derivation of its stress contour through the cyclical application of the NSR. The phrases 1-7 below the stress level numerals are the phrases used as test utterances in this study. They are positioned so that the numerals above these sentences represent both a step in the cyclical derivation of the stress contour of sentence A and the final stress contours of the test utterances.

2.2 Measurements

Mingograms were made of all repetitions of utterances and isolated words. Fundamental frequency, intensity, and duration were measured for all five readings.

2.2.1 Fundamental frequency: Fundamental frequency was measured at the temporal midpoint of the vowel of the words in the test utterances. Care was taken to produce an artifact free representation of \( F_0 \). In the preliminary work for this investigation, two sets of data had to be partially rejected because it proved impossible to obtain reliable \( F_0 \) information. The voices of the informants were simply acoustically incompatible with a satisfactory mingograph analysis. The informants also had many mispronunciations associated with emphasis and enumeration intonation. That part of the data that was not rejected showed qualitative similarity to the data represented below.

2.2.2 Intensity: Intensity was measured at the temporal midpoint of the vowel. An integration time of 25 msec was selected. The delay due to filter inertia required a temporal compensation of approximately 20 msec which corresponded to 1 mm in the sampling of the intensity curve.

2.2.3 Duration: Duration was measured for each stressed word from the release of its initial consonant to the end of its voiced segment.
## SPEECH MATERIALS

<table>
<thead>
<tr>
<th>SENTENCE A</th>
<th>He kicked the cop who thought that Dick should have bought the belt for his big pants</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYCLE 1</td>
<td>1 1 1 1 1 1 1 2 1</td>
</tr>
<tr>
<td>PHRASE 1</td>
<td></td>
</tr>
<tr>
<td>CYCLE 2</td>
<td>1 1 1 1 1 1 2 3 1</td>
</tr>
<tr>
<td>PHRASE 2</td>
<td></td>
</tr>
<tr>
<td>CYCLE 3</td>
<td>1 1 1 1 1 2 3 4 1</td>
</tr>
<tr>
<td>PHRASE 3</td>
<td></td>
</tr>
<tr>
<td>CYCLE 4</td>
<td>1 1 1 1 2 3 4 5 1</td>
</tr>
<tr>
<td>PHRASE 4</td>
<td></td>
</tr>
<tr>
<td>CYCLE 5</td>
<td>1 1 1 2 3 4 5 6 1</td>
</tr>
<tr>
<td>PHRASE 5</td>
<td></td>
</tr>
<tr>
<td>CYCLE 6</td>
<td>1 2 3 4 5 6 7 1</td>
</tr>
<tr>
<td>PHRASE 6</td>
<td></td>
</tr>
<tr>
<td>CYCLE 7</td>
<td>2 3 4 5 6 7 8 1</td>
</tr>
<tr>
<td>PHRASE 7=SENTENCE A</td>
<td>He kicked the cop who thought that Dick should have bought the belt for his big pants</td>
</tr>
</tbody>
</table>

**Fig. I-C-1.** Speech material used in this study presented in terms of its derivation. Sentence A and the phrases 1-7 are positioned so that the stress level numerals represent both a step in the cyclical derivation of the stress contour of sentence A and the final stress contours of phrases 1-7 which were used as test utterances.
3. Results

3.1 Fundamental frequency

The measured $F_o$ values were plotted as a function of the words in the test phrases. This is meant to throw light on the $F_o$ contour in each sentence and to facilitate discovery of any meaningful regularity in these contours. Fig. I-C-2 is a summary of these $F_o$ contours for all the test phrases considering only the stressed words. As each phrase is read from left to right, $F_o$ decreases in value between stress level 2 at the beginning and 1 at the end. Of particular interest here is that 1 has consistently the lowest value. This phenomenon has often been observed and discussed in the literature (cf. Lieberman 1966). It appears that as each expansion of the syntax was executed, the value of $F_o$ at the beginning of the phrase was raised. Therefore, the point of termination being around 100 Hz for all the phrases, the initial value and range of $F_o$ increases with increasing length (number of stresses) in the phrase. There is a higher rate of reduction in the range of stress levels, starting from 2 down to about 4. In other words, the rate of reduction is higher towards the beginning of every utterance than towards the end. The unstressed words in the phrases show an $F_o$ contour which is somewhat similar to that of the stressed words. The significant difference being that if the level of an initial unstressed word (which is always very low) is disregarded, the range of $F_o$ values is often somewhat smaller. For a full account of this data, see McAllister, forthcoming.

3.2 Intensity

As for $F_o$ the measured intensity values were plotted as a function of the words in the test phrases. This is intended to elucidate the intensity contours for each phrase and to make meaningful regularities in these contours more obvious. Fig. I-C-3 shows a summary of these intensity contours considering only the stressed words. It is obvious that these curves are similar in some important ways to the $F_o$ data. The descent of intensity value between stress level 2 at the beginning of each phrase and 1 at the end is not unlike the $F_o$ patterns although the rate of reduction is quite different. Note here again that stress 1 corresponds to the lowest intensity level. There seems to be here also, an approximate common terminal intensity level for nearly all the phrases. Unlike the pattern for $F_o$ where the stressed and unstressed curves gradually came together towards the end of the sentence, the two intensity curves seem to have no common values.
Fig. I-C-2. Summary of the $F_o$ contours for all the test phrases with $F_o$ plotted as a function of the stressed words in the phrase.
Fig. I-C-3. Summary of the intensity contours for all the test phrases with the intensity values plotted as a function of the stressed words in all the phrases.
3.3 Duration

The results of the duration measurements are presented in a way which is primarily intended to throw light on the correlation between the stress level output of the NSR and the duration parameter. Duration in milliseconds from consonant release to the end of the voiced segment for each test word has been plotted as a function of stress levels. Fig. I-C-4 is typical of the data plotted in this way. The word 'belt' was chosen as a representative example and all the other test words except 'pants' showed remarkable similarity to the data in Fig. I-C-4. This data shows some definite tendencies. Every test word (except 'pants' which always has stress 1) shows a relatively large difference in length of measured segments when the word has stress level 1 compared to the length when the word has level 2. The measured segments of the words were markedly longer with level 1 than with level 2. A central observation in the duration results is that there is relatively little difference in the length of the measured segments when the word has level 2 compared to when the word has any other level lower than 2. The word 'pants', which has stress level 1 in all the utterances (see Fig. I-C-5) shows somewhat longer measured segments when it is pronounced isolated compared to the occurrences of 'pants' in context. The difference between level 1 isolated and level 1 in context is, however, significantly smaller than the difference in all the other test words between the length of measured segments in words with level 1, obtained by isolated pronunciation, and words with levels 2 or lower. Similar results have been obtained for Swedish (Lindblom, 1970).

4. Discussion

4.1 Fundamental frequency

There seems to be no linear correlation between the output of the NSR and the $F_0$ parameter in the speech wave. When first confronted with the results shown in Fig. I-C-2, a correlation might seem evident. As the stress level decreases, i.e. numbers get larger, the $F_0$ curve slopes downward, i.e. the $F_0$ values get lower. Upon closer examination of the NSR, however, and of Chomsky and Halle's view of stress, this conclusion about the correlation should probably be judged as illusory. Chomsky and Halle's use of the numbers which represent stress level in Chomsky and Halle (1968) would indicate a greater degree of perceptual prominence in connection with
Fig. 1-C-4. Duration in milliseconds from consonant release to the end of the voiced segment for the word 'bought' plotted as a function of stress level outputs of the NSR. Data for all the other non-final stressed words in the test utterances showed remarkable similarity to 'bought'.
Fig. I-C-5. Duration in milliseconds from consonant release to the end of the voiced segment for the word 'pants' plotted as a function of the stress level of the preceding syllable.
higher level of stress. It is clear that when we consider the NSR, prominence does not bear a 1-1 relation with $F_o$. The point of the NSR is that it unfailingly gives stress priority to the right-most constituent in the phrase. Therefore in these strictly right-branching structures, where the NSR is cyclically applicable, the right-most constituent should always have the highest of whatever the stress levels represent. The results of the $F_o$ measurements show that stress level 1 consistently represents the lowest $F_o$. This discrepancy between the stress level contours which are outputs of the NSR and the $F_o$ parameter would lead us to believe that there must be a more effective way to describe $F_o$ patterns in right-branching structures.

The results presented in Fig. I-C-2 show evident regularities in the $F_o$ patterns. According to Chomsky and Halle, the central criteria for the operation of the NSR are bracketing of the surface structure and the labeling of the syntactic categories within the brackets. With this information, the NSR can then generate the proper stress contours. The results show, however, that the output of the NSR has no direct correlation in the $F_o$ patterns for the stressed words shown in Fig. I-C-2. The $F_o$ contours for the unstressed words also show obvious regularities and these regularities seem to have a definite relation to the contours for the stressed words. The description of the $F_o$ values for the unstressed words in these constructions then, can be made in terms of the relationship between the stressed and unstressed curves with the stressed functioning as a reference.

4.2 Intensity

Some of the discussion concerning the results of the intensity measurements will be similar to that of the $F_o$ results. As was the case for the $F_o$ curves in Fig. I-C-2, it appears difficult to discover meaningful correlation between the intensity patterns and the stress level contours which are the outputs of the NSR. The same, somewhat misleading tendency for agreement between decending stress levels and lowered intensity values is apparent in Fig. I-C-3. The validity of the stress level numerals as indicators of degree of intensity is, however, as for $F_o$, weakened by the fact that the level 1 has consistently a relatively low value in Fig. I-C-3. We must then establish that if the higher stress level numerals are to indicate a higher degree of any characteristic, then this characteristic can hardly be intensity. It is obvious that the intensity envelopes show some similarities for
the $F_o$ patterns in Fig. I-C-2. The general tendency for gradual reduction of value from high, left-most stress level 2 to a low approximate common value at the right-most level 1 is evident. Except for the shorter utterances, the intensity curves are quite similar in terms of rate of reduction (ca 2.5 dB/stress). This may lead us to believe that there is a general tendency for the intensity levels to be grouped along the same axis, and as each expansion of the syntax was executed, the intensity values for the added stressed words were simply added to the other at the appropriate place on the axis, i.e. higher and to the left of the others. The slope of this axis may be more clearly stated if we first take into account the observation that vowel quality has an effect in intensity levels. From Lehiste (1970), pp. 120-123, we know that close vowels have somewhat lower inherent intensity than the open vowels. An adjustment of the intensity levels of the words in the utterances according to these facts would raise the close vowels' and lower the open vowels' intensity slightly.

As was the case for the unstressed word's $F_o$, the unstressed word's intensity levels appear to have a direct relationship to those of the stressed words. The contours for the unstressed words, with the possible exception of the relative pronoun 'that' could then be defined in terms of the contours for the stressed words. The data for the pronoun 'that' is probably too limited to lead to any suspicion that this tendency for a lower intensity level has any generality for this or any other relative pronoun.

### 4.3 Duration

In the results of the duration measurements, we see perhaps the most pronounced discrepancy in the correlation between the phonetic facts and the stress level contours which are outputs of the NSR. One could expect that if duration were the characteristic described with the stress level numerals, a lowering of stress level would entail a durational reduction. Fig. I-C-4 shows that this is the case for the stress level change from 1 to 2 but not for stress levels lower than 2. There appears to be two durational categories for the syllable nuclei in these test words. There is a long category which occurs when the words are pronounced in isolated position and a short category which occurs when the words are pronounced in context. This is the case for all the words in the utterances except 'pants'. The syllable nucleus in this word has a more constant duration value than those of any other words. As mentioned in the presentation of the results, the isolated
'pants' is slightly longer than the occurrences in context. The difference, however, is small compared to the other words. It is important to keep in mind that the word 'pants' which shows this relative invariability in duration of measured segments, always has the stress level 1. These utterances, then, could be said to show two durational levels instead of the maximum of eight as the NSR would lead us to expect. First the durational level 1 and then a durational level less than 1, let us say 2. The level 1 occurs at the right-most constituent and in isolated position. The level 2 elsewhere.

An important question here concerns the meaning of the fact that the right-most element in these structures has a durational value equal to that of its isolated value. This could mean that stress priority given to right hand constituents in the NSR holds for the duration parameter. One can hardly help feeling a certain hesitancy to accept this correlation between the NSR and the duration parameter at face value. The well-known tendency for utterance-final syllable to undergo lengthening should be borne in mind.

5. Summary of the discussion

Consideration of the facts presented thus far would seem to indicate that the output of the Nuclear Stress Rule does not show a simple correlation with the prosodic parameters of the speech wave. This would be adequate motivation for us at this point, to systematically evaluate this rule as a means of predicting acoustic parameters in the speech wave.

5.1 The NSR-based model

In this evaluation, an NSR-based model for the prediction of the F0 and intensity parameters in this study will be compared to an alternative model based on the number of stresses in a phrase and position of words in the phrase. The duration parameter will not be included in this comparison.

In an NSR-based model we must make use primarily of outputs of this rule as a basis for predictions. The stress contours which are outputs of the NSR are expressed in terms of numerals, the numeral 1 representing the greatest degree of some stress characteristic. A reasonable demand that may be placed on this rule is that it, through its stress contour outputs, show a correlation between the syntax (information the rule makes use of in its operation) and phonetic reality by predicting the parameters we have discussed in this paper. The capacity of the NSR to do so is illustrated in Figs. I-C-6 and I-C-7. It is immediately apparent in these figures that at
Fig. I-C-6. A summary of observed $F_0$ values for each stress level including data for all the stressed words isolated and in context. The line connects the median values for each stress level.
Fig. I-C-7. A summary of observed intensity values for each stress level including data for all the stressed words isolated and in context. The line connects the median values for each stress level.
least two problems are associated with the NSR-based model. First, there is no linear correlation between descending $F_o$ or intensity values and descending stress levels. The second problem apparent in Figs. I-C-6 and I-C-7 is the spread of the data. The NSR-based model not only does not account for the spread, but perhaps more important, does not account for the obvious regularities in this spread. Use of an NSR-based model for the $F_o$ and intensity patterns in this study would require considerable adjustments of the NSR outputs. These adjustments would be based on some of the same factors which an alternative model would make use of for prediction of $F_o$ and intensity such as the number of stresses and their position in the utterance. Thus the low $F_o$ level of terminal stress 1 would have to be included as a basic rule.

5.2 Alternative model based on phrase length and position

Examination of the figure material presented in this study leads to the conclusion that it is possible to formulate rules that have quite different characteristics than the NSR. Instead of the syntactic information used by the NSR, it is possible to derive the $F_o$ and intensity values for the words in the test phrases by considering (1) the length of the phrase and (2) the position of the individual word in the phrase. In the formulas presented below, position is numbered from left to right. The left-most position in the phrase has the position 1, the next position to the right 2, and so on. The derivation of the values for the unstressed words is also included in this quantitative treatment of the data.

The following formulas are suggestions for derivation of $F_o$ and intensity values for each word in the test phrases used in this study.

$$F_o = \frac{(45 \log_{10} n + 9)(1-k)}{p} + 90 \quad (1)$$

$n =$ number of stresses in phrase
$p =$ position
$k = .2$ for a succeeding unstressed word
$k = 0$ for stressed words
Intensity \( I_o = n - 15 - 2.5(p-1) - 7k \) (dB) \( (2) \)

- \( n \) = number of stresses in phrase
- \( p \) = position
- \( k = 1 \) for succeeding unstressed word
- \( k = 0 \) for stressed words

In Figs. I-C-8 and I-C-9, observed values for \( F_o \) (Fig. I-C-8) and intensity (Fig. I-C-9) have been plotted as a function of values predicted with the formulas above. It appears that these formulas display a reasonable degree of adequacy in predicting these parameters for the phrases in this study. The fact that these predictions can be made with only two variables, position and number of stresses, which in any case would have to be incorporated into the adjustments of NSR outputs to obtain equally satisfactory predictory capacity leads us to the conclusion that the NSR is, in fact, of no use in these predictions.

6. Conclusions

It was implied in the introductory comments to the discussion of the results, the strength of the conclusions could be said to be diluted by the scope of the material and the approach to the problem in this study. While it seems necessary to qualify the content of this paper in this manner, it may be reasonable to assume that the conclusions shed some light on prosody in American English. This assumption raises the question of Chomsky and Halle's view of stress and the role of the phonetic representation expressed in Chomsky and Halle (1968). They feel that the stress contours which are outputs of the NSR, and which "are recorded by careful phoneticians,... constitute some sort of perceptual reality for those who know the language in question" (p. 25). They also state that "the speaker of English can determine the phonetic shape of an utterance on the basis of such rules as the Compound and Nuclear Stress Rules...." (p. 24).

The phonetic representation, then, can be viewed as a set of commands to the speech mechanisms about the speech signal that is to be pronounced.

---

* For a discussion which throws light on the validity of Chomsky and Halle's assumption about linguists ability to bring the stress contours "to a level of awareness", see: Lieberman, "On the acoustic basis of the perception of intonation by linguists", Word 21 (1965), pp. 40-54.
Fig. I-C-8. Observed values for $F_o$ plotted as a function of values predicted with the formula (1) section 5.2.
(Strangely enough, Chomsky and Halle argue that the phonetic representation need not have any relation to a "physical or acoustic reality", p. 25.) If rules provide outputs in terms of a phonetic representation with little correlation to the speech wave produced, we must conclude that the demands of the phonetic representation must be better defined, and that rules which give rise to this phonetic representation be accordingly formulated.

References:


