Stress and the Elastic Syllable: An Acoustic Method for Delineating Lexical Stress Patterns in Connected Speech*

Jane H. Gaitenby
Haskins Laboratories, New Haven, Conn.

ABSTRACT

Particular lexical stress patterns are common to the speech of native talkers of standard American English, but these patterns may be produced in a variety of prosodic ways. In this report, a technique is described for the retrieval of prosodic contours from the acoustic record (of a sentence as read separately by four individuals) that agree well with lexical stress patterns, as perceived in a pilot study.

INTRODUCTION

The main purpose of this report is to describe a method of deriving lexical stress patterns from the acoustic record of connected speech. Secondarily, it will be suggested that larger prosodic contours may be revealed by the same method. Prosodic data for one long sentence will be presented, and will be compared by talker, by prosodic parameter, and by summed parameter values in sequential syllables.

In this report, stress is defined as the property that endows sequential syllables with differentiating grades of acoustic prominence. The prosodic features that interact in signaling stress are: fundamental frequency—to be referred to below, with prosodic license, as "pitch"—duration, and intensity. Selected acoustic measurements of these three features comprise the data for the study. Spectral distribution, which is also generally acknowledged to be a stress cue, will not be explicitly referred to here.

The state of stress research can be summarized by noting that a great deal of what is known, and of what is known to be unknown, on the subject of stress

*This report is an expanded version of "The Elastic Syllable: An Acoustic View of the Stress-Intonation Link," a paper that was presented at the 88th meeting of the Acoustical Society of America, St. Louis, Mo., 4-8 November 1974. [J. Acoust. Soc. Amer. (1974), Suppl., 56, S32 (Abstract P5).]

Acknowledgment: Franklin S. Cooper introduced the author to the investigation of prosodic problems in English speech, such as the one described, and has provided guidance at just the right intervals. This is deeply appreciated. Warm thanks, too, to John M. Borst for his patient advice on instrumentation and measurement.

[HASKINS LABORATORIES: Status Report on Speech Research SR-41 (1975)]
today, was well described—under the heading of "Accent"—as early as 1934 by Carhart and Kenyon (1934) in the Guide to Pronunciation in the second edition of Webster's New International Dictionary.

[For further general background on stress research, the reader is referred to a concise review of the literature given by McClean and Tiffany (1973) in introductory paragraphs to their article. Also provided in the article are significant acoustic data and observations on effects of position, loudness, and rate on stress realization.]

Stress research is complicated by the fact that the three acoustic parameters acknowledged as cosignals to stress also apparently share in signaling another speech attribute, namely, "intonation." Intonation is thought by many to refer only to the perceptual phenomenon of pitch variation across an utterance, and the majority of intonation studies accordingly have been concentrated only on fundamental frequency contours. [Noteworthy exceptions are Denes (1959), Denes and Milton-Williams (1962), and Lieberman (1967).]

A further problem in investigating stress is that there are several types of stress that should be distinguished: lexical, semantic (under which we include contrastive and emphatic stress), and positional stress. These may co-occur in speech and thus confound analysis.

Another fact that makes stress description and analysis difficult is that stress perception is dynamic (as indeed is all speech perception), but acoustic displays of speech, such as spectrograms, immobilize the speech wave, leading to descriptions of the physical record that appear to deal with static events.

Published objective descriptions of stress have been fragmentary. If the corpus of speech examined in a study is relatively long, then the stress-signaling parameters described are probably few. Conversely, if two or three prosodic parameters are dealt with in detail, then the corpus itself is probably brief—consisting of nonsense syllables, single words, or extremely short sentences. (Furthermore, spontaneous natural speech is seldom used in stress experiments; text readings are preferred because they provide controlled verbal content.) In physiological research on stress, for which improved instrumentation and analytic techniques have been arduously developed in recent years, published reports have thus far, understandably, been confined to the behavior of only scattered portions of the vocal apparatus. Finally, very few accounts of stress experiments involve parallel data from more than one or two of the possible approaches to speech research, which may be physiological, acoustic, perceptual, and synthetic. [Lieberman (1967) is one of the exceptions.] Therefore, the analysis of stress remains partial and primitive, owing to the lack of multifaceted data on sizable stretches of natural connected speech.

[It may seem somewhat surprising that speech synthesis by rule is as good as it is, in view of the poverty of information available on stress. One reason for the high intelligibility of some versions of current synthetic speech must lie in the adequacy of the segmental rules used, including extremely good rules for duration. (Duration is the most tightly structured of the prosodic features in English, as will be illustrated below in natural speech data.) Aside from duration, considerable prosodic variation (elasticity within and across syllables) is permissible in the language. Mild prosodic (and phonetic) deviations from the norm, such as those heard in synthetic speech, may be heard as
dialects--to which most listeners can adjust themselves, as long as the variations are regular.]

Having expressed the need for more extensive investigations of stress, and having produced and noted the preceding caveats, the scope of this paper is nevertheless limited, in that it deals only with data from the acoustic plane. In this paper we describe an approach to the characterization of lexical stress by way of a measurement and display technique that delineates acoustic patterns corresponding closely to intrinsic stress patterns.

The present paper offers a new look at prosodic measurements that we made between 1958 and 1960. The material pertaining to the speech sample used, the method of measurement, and the measurement units themselves therefore date from that time, when the purposes of the experiment were to produce an acoustic description of running speech and to find correlates of stress in that acoustic record. It was thought sufficient at that time to characterize acoustic stress in a relative manner, by merely noting whether the combination of pitch, duration, and intensity parameters in a syllable were higher or lower than the prosodic combinations in immediately adjacent syllables. In recent reexaminations of the same acoustic data, it has appeared that more informative stress patterns can be revealed by referring to the absolute prosodic measurements. It is this latter approach that will be presented, after the procedure used in the initial acquisition of the data has been described.

I. DATA ACQUISITION

A. Initial Assumptions

Two assumptions were made at the outset of the original experiment:

1. Peak pitch, peak intensity, and total duration of voicing in a syllable are sufficient data to characterize syllable stress (relative to adjacent syllables).

2. Syllabic acoustic data for these three prosodic parameters can be combined to produce total prosodic (stress) value of a syllable. (The three parameters share the attribute of signaling stress perceptually; it is therefore reasonable to assume that acoustic parameters combine to signal stress.)

B. The Corpus

The speech material used consisted of readings of a text (about 500 words long) that was created from a selection of high-frequency English vocabulary, including polysyllables as well as monosyllables (Dewey, 1923; Thorndike and Lorge, 1944). Several of the polysyllables were intentionally repeated, at least once in the script, in contrasting locations, and in differing grammatical roles where that was possible, e.g., "official" was used as a noun in one sentence, and as an adjective in another. Words in which stress patterns change with grammatical, semantic, or positional usage (such as "transport," "invalid," "absolute") were not used.

The form and content of the text was like a dull governmental announcement (high-frequency polysyllables from word counts of printed matter suggest that semantic field) and most of the sentences were long. Unemphatic readings at normally fast speaking rates were required, and it was assumed that the long and
Interesting sentences would contribute to those effects. It was also anticipated that the intrinsic stress patterns of the polysyllables would be very reduced in such a context; therefore, any evidence of acoustic correlations with lexical stress patterns might be considered basic stress cues.

The text was read, casually and rapidly, by three men and one woman from the laboratory staff. Each person was recorded at a tape speed of 15 ips under standard sound-proofed room conditions. The talkers were native to the United States and spoke "eastern educated speech," although their maturational years were spent in various parts of the country. Their ages ranged from 30 to 42 years.

Three focal sentences, containing among them two or more instances of certain polysyllables, were excised from each person's tape recording and were then measured by the means to be described. The shortest of the sentences (29 syllables long) will be discussed in this paper.

C. Measurement Method

The speech waveform and hill-and-dale trace of the pitch level were recorded by dual-beam cathode ray tube on 35-mm film at 7.2 ips. (The pitch voltages were taken from a conventional Vocoder.) The pitch values were calibrated against 125-msec tape-recorded sequences of 80-, 90-, and 120-Hz pure tones that had been spliced into the source audio tapes. Measurements of peak pitch and total duration of voicing were made by reference both to the film and to wide and narrow band spectrograms; the amplitude curves above the wide band displays were used for the peak intensity measurements.

The syllable boundaries were marked consistently and in corresponding positions on the film and spectrograms, for the four versions of the sentence, with word boundaries preserved because word stress patterns were to be compared.

D. Measurement Units

It will be seen (Figure 1) that the units of measurement are not conventional, although the pitch and duration data shown can be converted readily to traditional units, as will be described. The intensity data shown will also be explained.

Three constraints were taken into consideration before the decision was made on how the acoustic measurements might best be examined and presented as prosodic data:

1. There were limitations on the precision of measurement, imposed by the small size of the acoustic displays employed. For example, the resolution of the pitch trace (on film) permitted measurements no finer than in approximately 4-Hz units.

2. The numerical ranges of all three parameters had to be compatible in magnitude so that the parameters could be displayed and compared in parallel on a common grid.

3. Weighting of the parameters seemed desirable. [Bolinger (1958) had presented evidence for the primacy of pitch over duration as a
<table>
<thead>
<tr>
<th>TALKER</th>
<th>Pitch Duration (ms)</th>
<th>Duration (ms)</th>
<th>Intensity (ad)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>12.8</td>
<td>15.6</td>
<td>7.1</td>
</tr>
<tr>
<td>S</td>
<td>11.9</td>
<td>12.4</td>
<td>7.3</td>
</tr>
<tr>
<td>L</td>
<td>11.9</td>
<td>12.4</td>
<td>7.3</td>
</tr>
</tbody>
</table>

Figure 1
stress cue, and Fry (1958) had shown that duration was a stronger stress cue than intensity (in single words, at least). It seemed reasonable, then, to approximate the apparent hierarchy of cues in the graphic display of the prosodic measurements.

A practical method of working within these constraints was to estimate the likely ranges of the three parameters to be found in the four readings, and then to scale the separate parameters to represent the stress cue primacy of pitch over duration, and duration over intensity. To do this, acoustic measurements were converted to representative "prosodic units."

For pitch, the lower limit of the range was set at 60 Hz. [All syllables in which the pitch peak registered 60 Hz or below are called "0" (zero) in the Figure 1 data because very low pitch values are usually accompanied by low intensity levels, which are normally below the threshold of hearing.] An upper limit of about 200 Hz was assumed on the basis of preliminary inspections of the acoustic record. The resulting range of 140 Hz (60-200 Hz) was measured in 4-Hz units, producing a range of 35 "prosodic units." A range of 35 steps seemed sufficient for the display of syllable peak pitch contours.

The durational range was estimated at 20 to 400 msec, for the shortest to the longest syllables, voiced portions only. It was appropriate to measure duration in 20-msec units, which produced 20 prosodic units as equivalents to the anticipated durational range.

Syllable peak intensity measurement was made from a logarithmic plot of the rms voice amplitude in dB (the amplitude curve displayed on the Kay Sonagraph spectrogram). Maximum and minimum intensity values were found for each talker (i.e., a personal vocal intensity range). This range was divided into 14 equal linear steps—the smallest practical number of divisions. These were called the 14 prosodic units of intensity. Consequently, the prosodic unit of intensity may differ somewhat from talker to talker, but it is consistent across the utterance for each individual.

In short, to make weighted graphic comparisons, we measured the parameters in prosodic units that represented actual measurements on each parameter, but the number of prosodic units available to the display of each separate parameter was apportioned to suggest the rank ordering of the stress cues. Thus:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range of Prosodic Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch</td>
<td>35</td>
</tr>
<tr>
<td>Duration</td>
<td>20</td>
</tr>
<tr>
<td>Intensity</td>
<td>14</td>
</tr>
</tbody>
</table>

It must be emphasized that the values of the prosodic units for different parameters have been selected both as a weighting device and for graphic convenience. Actual stress equivalence is NOT implied between, for example, 10 prosodic units of pitch and 10 prosodic units of duration or intensity, although, for the purpose of the analysis to follow, they will be treated as equivalent.
E. Data

In Figure 1 the acoustic data in prosodic units are presented for each successive syllable of the sentence, "An official, that is a department head, hopes that you will understand what several of the officers' comments mean." The prosodic units shown there can be converted, if desired, back to traditional measurement units as follows.

For pitch in Hz, multiply the number of prosodic units given for a syllable on a pitch row by 4, and add 60. [For instance, for Talker R, first syllable, Pitch = 12. (12 X 4) + 60 = 108 Hz.]

For duration in msec, multiply the prosodic units given on a duration row for a syllable by 20.

The intensity units shown are relative within the speech of each particular talker. (The highest intensity measurement possible in any of the sentences was 14 prosodic units. In this sentence, the highest intensity value happens to be 12.)

When inspecting the data, the reader must bear in mind that the measurements refer to peak pitch in each syllable, to total duration of syllable voicing (which includes voicing in consonants as well as in vowels), and peak intensity in the syllable. The row of syllable Total values will be referred to in Figures 4 and 5 and can be ignored for the present.

II. ANALYSIS

A. Single Parameters, Compared

The data presented in Figure 1 are exploited in various graphic ways in Figures 2-5. In Figure 2, the prosodic unit data for the last portion of the sentence ("...hopes that you will understand what several of the officers' comments mean.") are shown by individual talker. The syllabic data nodes for each parameter, indicated by small circles, have been connected by (distinctive) lines in order to produce comparable prosodic contours across the utterance. Individual speaker differences in contour shapes and ranges of the trio of parameters are immediately visible. There are also resemblances across the speakers, notably in duration, as was expected (Caftanby, 1965).

It can also be seen that the pitch contour is closely paralleled by the intensity contour (or vice versa) in the records of Talkers R and L, but these contours are less clearly related for Talkers S and G. There are other differences to note: Talker R's intensity range is quite narrow, and R also tends to use more or less equal durations in some sequences. Talker S shows considerable fluctuation in all parameters. Talker G (the female in the group, with atypically low vocal register for a female) produces pitch excursions that are more extreme, and generally higher, than those of the other talkers. G's prepausal durations are also the longest. In Talker L's record, in contrast to the others, a clear falling trend can be seen in the pitch and intensity contours (with final higher peaks).

The main point that Figure 2 is intended to illustrate is that readings of the same verbal material by several speakers produce prosodic trio contours for each talker that look far from identical from talker to talker.
PITCH vs DURATION vs INTENSITY DATA, BY INDIVIDUAL TALKER
(LAST 18 SYLLABLES OF SENTENCE)

TALKER R

TALKER S

DURATION

INTENSITY

PITCH

PROSODIC UNITS

TALKER G

● = PAUSE

TALKER L

hopes
that
you
will
understand
what
several
of
the
officers'
comments
mean.

Figure 2
In Figure 3, cross-speaker comparisons are made by parameter for all 29 syllables of the sentence. We now focus on the generalized contours produced by the four talkers' versions of pitch, duration, and intensity. The talkers therefore have not been identified in these graphs (but they can be, by reference to the preceding figures).

Extreme similarity in relative duration is now obvious; there are few deviations from the essentially single pattern produced by the four individual duration contours. The talkers plainly conform in temporal organization of the sentence, although their individual speaking rates may differ. [Fundamental determinants of relative syllable duration are the number and kind of phonemes in the syllables, determined by the particular vocabulary used in an utterance. One might then assume that the common verbal material accounts entirely for the profound cross-speaker durational regularities, but preliminary studies we have made (unpublished) indicate that non-native talkers of American English produce different durational patterns from those of natives (unless their English rhythm is so "good" that it cannot be distinguished from that of native speakers). The cross-language durational problem also involves the consideration of languages in which there are phonemic length contrasts (see Peterson and Lehiste, 1960; Lehiste, 1970) and native versus non-native stress realization. However, we shall not pursue the matter here.]

Definite similarities in the four intensity contours also appear in Figure 3, as well as some general agreement (very strong in the initial phrases despite later occasionally contradictory slopes) in the overall pitch pattern. Note the four distinct pitch registers visible in the syllables of the appositive phrase, "...that is a department head,..."

The contours of all three parameters show fairly good agreement in rising and falling with the lexical stress patterns of the polysyllabic words, "official," "department," and "officers." A possible exception, at first glance, is the inherently low-stressed final syllable of "official" [ˈɔfɪʃl], in which the duration of voicing rises above that of the preceding syllable [I], seeming to indicate increased stress and/or different phonetic content. It is clear that [I] and [ˈɔ] are different phonetically, and that [I] is intrinsically brief. However, [ˈɔ] is prepausal, phrase final, and precedes an embedded parenthetical clause, all of which necessarily involve extended duration (Klatt, 1974). The slight rise in duration in this case is thus, at least in part, a conditioned effect. (Unless a rise in prepausal duration is very substantial—on the order of more than twice the normal length of a syllable of the given phonetic type in prepausal position—a stress increase is probably not indicated by a rise in duration alone.) Note here that all talkers' pitch and intensity peaks fall in [ˈɔ], counteracting the rise in duration.

One syllable in which greatly extended duration does appear to be the major signal to increased stress is the final syllable of "understand" [ənˈdænd]. This is an intrinsically long syllable in number of voiced phonemes, and it is also phrase final. Talker G paused after this syllable; the other talkers did not produce silence here. At least two of the other talkers, however, also appear to have used length as the prime cue to the stress rise, and they may have simultaneously produced a "pseudo-pause" [a term and concept from Coker, Umeda, and Bowman (1973)] at that syntactic break by means of the prolonged syllable duration.
Note in Figure 3 that no single prosodic parameter is a reliable acoustic cue to lexical stress for all of the four talkers. That is to say, no one parameter shows unanimous rises and falls corresponding to the inherent stress patterns of the words of more than one syllable.

B. Summed Parameters, Compared

It is widely known that perceived stress depends on the combined effects of the prosodic properties within a syllable (in relation to those of adjacent syllables). It is therefore reasonable, on the acoustic plane, to sum the values of the separate parameters within each syllable, and to compare and examine the prosodic contours so produced. Summing the values of the trio of parameters is permissible because the measurements have been converted to common prosodic units.

Figure 4 shows the contours resulting from plots of the parameter totals. (The data points shown here were taken from the rows called "Total" in Figure 1.) Despite the individual differences that showed up in the single parameter contours in Figures 2 and 3, here the four talkers' separate versions of the sentence are generally alike in overall prosodic pattern. Although minor conflicts among the talkers in syllable slope are observable in these contours, most conflicts are probably due to slight differences in the semantic-syntactic interpretation of the verbal material and to idiolectal variations. Note that the lexical stresses of the polysyllabic words, however, are reflected by all of the talkers by appropriate rises and falls—in all but four instances (of which one, the [f] in "official," has been discussed). In the second syllable of "several" (pronounced [vr[f]] by all of the talkers), one person produced a very small contour rise (contrary to the expected pattern of the word), and two of the four talkers each produced a decided rise on the second syllable of "comments." These conflicting slopes may be artifacts of the syllabification procedure used in these words, i.e., after the first vowel. If the syllabification had followed the articulation and perception more realistically, some portion of the voiced consonant after each vowel of the first syllable would have been included in the respective first syllable of each word, thus altering at least the value of the first syllable's duration in an upward direction. (This does not explain, it must be admitted, why the other two talkers nevertheless produced the "correct" falling slopes for the second syllable.) In the case of "comments," it may be significant to the contour that the vowel in the lexically less-stressed second syllable is relatively full-grade, and furthermore, that this less-stressed syllable is penultimate in the utterance, which is to say that it is likely to have received conditioned lengthening in that location.

There are other possible explanations for the rises found in the slopes of some of the unstressed syllables. The most obvious is that vowel color was (intentionally) omitted as a prosodic parameter in this experiment, although it is a fact that no English syllable containing either a schwa or a syllabic consonant is lexically stressed. A consideration of vowel color would thus mark the syllabic "f" syllables as unstressed.

Another reason for the occasional slope discrepancies may lie in the weighting method itself, which can easily be modified. In particular, the conditioned effect of position on syllable duration, alluded to previously, might be compensated for in a revision of the parameter weighting.
A further explanation presents itself if it is understood that lexical stress is based not only on the prosodic qualities of the syllables interior to a word, but also on the prosodic relationships between the word and its context, especially the relationships with adjacent syllables. Thus, if syllable A, e.g., [(k)ə], of a two syllable word A + B, e.g., [(k)əmən(ə)], is preceded by a weak syllable in the previous word, then the stress of syllable A will be enhanced—and if syllable A is preceded by a weak syllable and followed by a comparatively small rise in syllable B, then the large positive rise leading up to A will probably override the small negative stress effect on A provided by the shallow rise from A to B, and A will be heard as more stressed than B. And, if rising syllable B, in the circumstance just described, is followed by a substantial rise in the next word (as is the case in [min] following the aberrant rising slopes seen in [mən]), the stress value of B will be further diminished. (These apparent contextual effects seem entirely logical, and a pilot perceptual test points to their validity, but they remain to be tested rigorously.)

To summarize Figure 4, we have seen that the acoustic contours (produced by summing the parameters in weighted prosodic units) reflect lexical stress patterns in nearly all the cases; only 4 of the 64 syllables (or 6 percent of the slopes) in the words of two or more syllables were "wrong." It may be inferred then, that larger prosodic patterns—such as phrasal stress patterns—are also as satisfactorily represented in the acoustic contours so derived. In short, although the measurement technique that has been described is cumbersome by the manual methods that were employed, it has utility for stress retrieval on the acoustic plane, and it should be reasonably simple to automate, given preestablished syllable boundaries.

Although the essence of our approach has been given in the preceding figures, and particularly in Figure 4, an additional view is presented in Figure 5 (top contour, in heavy line) in order to show the average of the four talkers' contours (from Figure 4), which can be considered the basic stress profile for the sentence. In the heavy contour all of the lexical patterns are correct in general shape, with the exception of syllable two in "comments," already discussed. Note, for example, the correct contrasts in the pattern of the word "official" versus "officers," and compare the contour of "understand" (stress pattern: mid, low, high) with that of "officers" (high, low, mid).

A few further observations can be made on the basis of this generalized contour (which is also representative of the contours found in the two longer sentences that were examined). (1) Local peaks in the contour are the relatively stressed syllables, and local valleys are unstressed syllables. (2) Peaks with the deepest valleys (immediately adjacent) are the stressed syllables of words that are high in information in the utterance. (3) The stressed syllable at a peak is usually part of a content word. (4) Relatively flat valleys consisting of two (or more) syllables are likely to contain at least one function word. (5) Rising slopes of the contours appear to contain more significant semantic information than falling slopes. (6) When the local peaks are connected by lines to produce a supralexical prosodic contour, the peaks then produced are the stressed syllables of words of major semantic importance in the utterance (i.e., key words).

It should be mentioned that a variety of phoneticians who have examined these data report that the contours shown in Figures 4 and 5 (top) closely resemble their intuitive impressions of the expected intonation pattern for the
Figure 5

AVERAGED TOTAL UNITS FOR FOUR TALKERS
and (below that)
AVERAGED PITCH vs DURATION vs INTENSITY FOR FOUR TALKERS

Figure 5
sentence. As has been demonstrated, "total syllable contours" are produced by summing the prosodic contents of the successive syllables, and therefore represent duration and intensity prominence effects in addition to those of pitch, although pitch (it will be recalled) is more heavily weighted in the display.

The averaged separate parameters are shown in the lower portion of Figure 5. It can be seen that all three run nearly parallel courses, except that duration rises prepausally, which is its normal configuration. Because these contours represent averaged data, it is not surprising to observe highly cooperative tendencies in slope across the parameters, but such profound agreement in the prosodic trio throughout an utterance is apparently less common in the speech of individual talkers (as has been illustrated in Figure 2).

**CONCLUSION**

The merit in using the procedure described is that it quite dependably elicits reasonable lexical stress patterns from limited acoustic information on connected speech.

We have also used a version of this technique in estimating (by eye) stress relationships in the syllables of unknown utterances—in spectrograms—with helpful results. However, the technique has an important prerequisite: the extent of syllable voicing must be known (i.e., syllable boundaries must be pre-established) in order to proceed with prosodic measurement, summing of syllable values, and the construction of contours. Several tests of this stress retrieval method have also been initiated using an algorithm for automatic syllable segmentation (Mermelstein and Kuhn, 1974) as a point of departure, and the results are promising. (It must be noted that segmentation of lexical boundaries, as such, is not attempted.)

It was observed above that a very few syllables (for individuals) failed to produce the expected prosodic contours. Two of these aberrant cases involved an unstressed syllable—containing a syllabic "ə"—that was long in voicing, following a stressed syllable that contained a short vowel, e.g., "official" [ə-fəl-əl], "several" [ə-vər-əl]. We are mindful of the fact that not only the voiced portions of speech, but also the voiceless regions, contribute to stress effects, even though, for the purpose of this experiment, the voiced portions were assumed to be the significant prosodic domain. In perceptual tests of stress that are being run, pairs of sequential syllables from the described sentence, spoken by individual talkers, are presented in two stimulus types in order to compare the prosodic contributions of the voiced portions alone (in one test) with those of whole syllables (in another test), and to compare the results of both of these test varieties with the acoustic contours derived as shown here. A pilot test, employing only the voiced portions of syllable pairs, indicates strongly that the acoustic contours—derived as in Figure 4—do, in fact, reflect perceived lexical and phrasal stress patterns. (Over a hundred syllable pairs from utterances by three of the four talkers have thus far been presented to six listeners.)

In the Introduction, it was mentioned that a total of three very long sentences were examined acoustically, of which the one that has been described above was the shortest. It is worth noting that the polysyllables that appear only once in this shortest sentence appeared at least twice in the three-sentence
sample, and very similar lexical patterns--different only in degree--were elicited in the acoustic contours for each version of a given word, despite changed grammatical usage and/or sentence location.

REFERENCES


152