The Elastic Word†
Jane H. Gaitenby
Haskins Laboratories, New York City

Anyone who has listened carefully to speech or who has examined acoustic records of continuous speech is fully aware that words stretch, shrink or stretch-and-shrink on all dimensions. Specific speakers, specific stress, intonation, phonetic context, and sentence position, (among other things), all simultaneously affect the specific shape of a given word. Some of the changes that appear on the acoustic record, however, are not noticeable to the listener, because the changes are regular or conditioned. For example, the phrase, word or syllable just before a pause tends to be long. Most people are not aware of this as such, but if the phrase or word is shorter than usual, or longer than usual, whatever "usual" is, then the duration becomes significant and some aspect of the message has changed.

We are interested in finding "usual" phonetic and prosodic patterns in order to predict typical variations in word shape. Phonetics and prosody are inseparable in real speech, but this paper will deal only with prosody, and only with one aspect of that: duration. A speech description that deals with any single prosodic feature is not quite doomed from the start. It is only dim from the start, because a word is the product of multiple cues to its identification. A case in point is syllable prominence, where the combined cues of duration, intensity and fundamental frequency are significant. This makes it clear that research on any one prosodic feature is a chancy business.

3.1
In spite of this, one prosodic feature may be stronger than another in producing speech effects other than stress. This is the rationale for considering duration alone here. Any apparent regularities that arise may be strengthened when cross-parameter data are combined at some future time.

On the hand-outs are samples of normal duration. The words are in continuous speech, which is the natural habitat of words - as opposed to citation forms which have received much attention in the literature. The drawings represent tests of how duration is affected by individual speakers, by dialects, and by pauses and utterance length.

Hand-out One, "Segment durations for fast and slower speakers," illustrates real-time differences in speech length shown by five adult speakers. Three are male, D.H. and J.G. are female. These five were selected as representatives of fast, slow and medium rate talkers. All five have no other native language than American English, are college-educated, and are between the ages of 25 and 50. Their speech was recorded under identical conditions. The short text shown is part of a postscript to a script which contains sentences that were to be read casually. After reading the sentences, each speaker stated whether he considered himself a fast, medium or slow speaker. The readings of "I consider myself (a this or that kind of speaker)" are therefore quite conversational in manner.

The top half of Hand-out One shows the five versions of the phrase on a 5" to 1 sec. scale, as on the usual spectrogram. The total rate of Speaker A.A. is about twice as fast as the rate for R.B., on the fifth line down. (Speaker A.A. is considered a rapid speaker by his colleagues, and R.B. is thought of as a fairly slow speaker - so casual perceptual impressions of speed tally with this acoustic sample.) This suggests that the speakers chosen are good representatives of a normal range of
speech rate.

What you have before you is almost raw data. Short vertical lines delimit segments of the utterance which are easily measurable on spectrograms, such as stop closure, release and "vowel formants." Extra horizontal lines have been drawn on alternate blocks to emphasize the similarities and differences in the length of a given segment in a comparison across the speakers. Where there is a vertical line through the "k" in "consider," a brief frictional burst is indicated. If an "s" block is subdivided, the large part is friction, and the small part is silence.

Now look at the dotted lines running diagonally from the upper left to the center right. The dotted line beginning in the "n" of "consider" for A.A. connects that point of each speaker's phrase which is 25% of the total duration of his or her utterance. The next dotted line, connecting the 50% points, lies between the midpoint of "-er" and the first part of the "my" syllable for the five talkers. The 75% mark is always in the "se" part of "self." Please note that consonant transitions may occur fore and aft in the formant segments symbolized only as vowels here. The so-called vowel formants actually contain perceptually important parts of the adjacent consonants. These segments are acoustical ones which are visible on spectrograms; they are not the usual auditory segments. An attempt to split syllables into such microauditory slices by listening alone, would be a task out of Greek mythology.

At the lower half of the page is another view of the same data. The first vertical column compares the duration of each first segment, "aj." The second segment, "k," is compared in the second column, and so on. A.A.'s segments are always shorter than R.B.'s, and the rank order of duration, brief and fast at
the top of the columns, long and slow at the bottoms, is maintained in a general way. The absolute similarities in the "n," the top three "d" closures, the "m" and the friction of each "s" are noteworthy. The difference in the "aj" spoken by J.W., as compared with the other speakers implies that he used less stress than the others, and the tape recording confirms it. The same is true of speaker A.A. in the last two segments.

The speech sample displayed here is brief in the extreme, and the number of talkers is small. Four observations summarize this test:

1. The total duration of a rather short utterance may be expected to vary with the speaker. Here the fastest speaker covers the same ground as the slowest speaker in 50% of the time.

2. Wide variations in length may be expected on stressed syllables.

3. Most of the segments, in relation to the total utterance, have approximately the same ratio for all the speakers.

4. Voiced consonants may be exceptions to the apparent segment-to-utterance ratio. Here they are very close in absolute duration for all five talkers.

Improvements and refinements for the next test design present themselves on the basis of the data above. For one thing, the number of speakers sampled should be increased to confirm the observations, and to confirm that the range in speaking rate represents a standard American range. For another, the reliability of speech rates deduced from brief samples should be checked in another type of utterance.

Hand-out Two, entitled "Cross-dialect comparison of duration," embodies the revisions just mentioned. In this test the question, "Can't you talk to Dave tomorrow?" was read by each person as one of several sentences in a script. The script instruction was to read naturally and casually, as if in conversa-
tion. The top section of Hand-out Two shows the same five speakers as in Hand-out One, and in the same order. Notice that each of the five is still relatively fast, medium or slow, as he was before. The speech rates indicated in the first test persist here for all the subjects. The ratio of segment length to utterance length continues to look constant for all the speakers, because the dotted lines fall in the same regions for all.

Eleven new speakers are shown below the original five. They read the same script, but were recorded at Stanford University, through the kindness of Prof. Dorothy Huntington, Michael May and associates there in the Department of Speech Pathology and Audiology. I hasten to add that these subjects, like the former ones, were completely normal American talkers. None of these new subjects was aware of the purpose of the experiment either. All are college graduates, twenty to thirty years old, and are students in the aforementioned department. All are female except three, M.M., J.W. and J.C. Unlike the first five, these new subjects were selected because they came from various geographical areas of the country. Their dialects, as heard, do not represent their widespread places of origin and linguistic development.

Now let's look at Hand-out Two again. The Stanford speakers (on the lower half of the page) are all one-quarter of the way through the utterance at the beginning of the word "talk." They are half-way through as the word "Dave" begins, and all have reached the 75% point early in the word "tomorrow." The Stanford subjects' ratios are in marked agreement with those of the Haskins subjects.

It should be recalled that the Haskins group had been selected because their speech rates seemed to represent fast and slow talkers. The Stanford subjects, on the other hand, were chosen particularly to represent several dialects. (We have listed the dialects crudely, by state only.) The points to note here
are that individual differences may be more important than dialectal differences in affecting speech rate; and, no matter what the speech rate or dialect may be, a segment ratio appears common to all the speakers. Parenthetically note that the range of total duration is close for both groups, although the Stanford median is on the long end of the Haskins sample.

(Incidentally, you may have noticed that the "n" of "can't," the "v" of "Dave" and the "r" of tomorrow" are set off in the drawing by broken vertical lines. These show that the acoustic margins were indefinite on the spectrograms.)

The segment ratios were checked statistically. Every segment was measured to the nearest 5 msec, and these measurements were converted to percentages of total utterance time. Then the segment percentages were compared across all sixteen speakers. The results were that nine-tenths of all the segments varied less than 5.3% for any speaker. For example, the " " segment of "talk" ranged form a low of 6.9% of total duration for R.B. of Massachusetts, to a high of 11.2% of the total for J.C. of Mississippi. This a range of 4.3%. Half of all the segments varied less than 3% from speaker to speaker. The greatest range, 7%, occurred only in the two multi-phoneme segments, the "ejv" of "Dave" and the "row" of "tomorrow." They are not strictly comparable segments on this account.

Great similarity in relative duration despite dialect or idiolect is evidence for one durational regularity in American speech. This should help in the study of intonation and tempo, if indeed tempo and intonation are separable. It should be remembered, though, that if larger segments such as words are compared, durational ranges will be larger, because sub-syllable differences are cumulative.

In general, slow speakers tend to be slow all along the line in their acoustic segments, but less slow in critical one-
stroke gestures such a stop releases. The time differences between these fast and slow speakers in pure speech are greatest in vowels in stressed or pre-pause positions.

This augments rather than contradicts a study on the determinants of speech rate by Frieda Goldman-Eisler in 1956\textsuperscript{1}. She stated that it is the duration of pauses that differentiate slow speakers from fast ones, rather than the speed of actual speech movements. Without reference to pause length or frequency, we find that speakers perceived as slow produce relatively long syllables, whereas fast speakers use short syllables. However, Dr. Goldman-Eisler compared the average syllable length for her eight subjects, and they spoke spontaneously. Both her conditions and her method differed from ours. We used scripts, without sacrificing speech naturalism, we believe, and kept the words to be compared identical for all speakers. Dr. Goldman-Eisler compared average syllable length, measured in both short and long utterances, probably. Syllables are long in very short word groups. In long utterances, syllables in non-final words tend to shorten toward an optimal minimum. Utterance length differences may explain Goldman-Eisler's varied findings on pure speech rate.

Syllable duration in short versus long sentences is demonstrated on Hand-out Three. The sentence on each line is expanded by the addition of a phrase on each lower line. J.W., a Haskins subject, read all of these samples among a number of others, in which these were embedded out of succession. The style was conversational.

Compare the first line, "Why don't you?" with the same three words in Line II, "Why don't you get tickets?". Many segments in Line II have shrunk, so that the overall phrase "Why don't you?" is 44% shorter than it was in pre-pause position in Line I. In Lines III, IV and V the phrase remains about as it was in Line II, despite the lengthened sentences. This suggests
that an approximate minimum duration was reached in Line II where it was not pre-pausal.

The next phrase, "get tickets" is pre-pausal in Line II, but follows three syllables. "Get tickets" reduces by only 20% in Line III where it is followed by a phrase. The reduction between the pre-pause position in Line II and the pre-phrase position in Line II is much less for "get tickets" than it was for "Why don't you" in the same kind of shift. The lesser reduction may be due to the fact that "get tickets" was not a solitary utterance to begin with in Line II, but a part of a longer utterance. Additional factors may be the different phonetic composition of these syllables, stress difference, and the fact that "tickets" may be less flexible in length than a monosyllable. In Lines IV and V, "get tickets" remains at about the same length as in Line III. The phrase, "for tomorrow," can be compared similarly.

Two basic rates of speech for any speaker are apparent, conditioned in a major way by the presence or absence of a pause immediately following the phrase. (This applies to a single talker using the same emotional tone in all cases.) The speaker appears to program his articulation rate as normally fast unless he is about to pause. This agrees with the relevant observations on Hungarian speech made by Fonagy and Magdics in 1960, though our tests were very different.

The phenomenon of pre-pausal word lengthening has been taken for granted in the past, and has not been considered as a factor in speech rate. (We are also inclined to include final syllabification of otherwise mute "e's" as in French, and even particle addition as in Chinese under the heading of pre-pausal lengthening.) If stretching the pre-pausal word is a universal tendency, then a physiological factor might be looked to as a cause of lengthening in this position. Conditioned pre-pause
word duration would account, at least in part, for the relative slowness common to short sentences and to citation words as well.

We can mention only in passing that the other possible factors which influence word duration are numerous, and frequently simultaneous. Some are physiological, others linguistic, and still others are specific to the individual talker and the occasion.

To summarize the conclusions suggested by the data we have presented: Firstly, individual variation in syllable rate is great, and a broad range of rates can probably be found in any American dialect. Secondly, a common ratio of segment-to-utterance length appears in all the dialects of American English we have sampled. Finally, we have observed that words immediately preceding a pause expand in utterances of all lengths. This leads us to the view that: a short utterance, which is normally slow, is a reflection of extended syllable duration indicating pre-pause anticipation. A long utterance, normally fast except at the end, reflects short syllable duration programmed by the speaker because no pause is imminent, plus final extended duration as the speaker anticipates an immediate pause.

Footnotes


The research described in this paper was supported by contract with the Prosthetic and Sensory Aids Service, Department of Medicine and Surgery, Veterans Administration.

References


SEGMENT DURATIONS FOR FAST AND SLOWER SPEAKERS

CUMULATIVE DIFFERENCES (ENTIRE PHRASE):

I CONSIDER MY SELF

SCALE:

1 INCH = 0.2 SEC.
(200 MSEC.)

SEGMENTAL DIFFERENCES:

"THE ELASTIC WORD" (SHAE GAITEY)

HAND-OUT #1.
CROSS-DIALECT COMPARISON OF DURATION

Script for both groups: "Can't you talk to Dave tomorrow?"

STANFORD U. TALKERS

"THE ELASTIC WORD" (JANE GASTANBY)
SEGMENT DURATION COMPARISON
IN LONGER AND LONGER UTTERANCES

ONE SPEAKER ONLY: J.W.

I was a young 
teen 

II was a 

III was a 

IV was a 

V was a 

PHRASE DURATION IN MSEC., AND AS PERCENTAGE OF SAME PHRASE IN PRE-PAUSE POSITION

<table>
<thead>
<tr>
<th>Why don't you</th>
<th>get tickets</th>
<th>for tomorrow</th>
<th>night (v, night's)</th>
<th>program</th>
</tr>
</thead>
<tbody>
<tr>
<td>msec.</td>
<td>percentage</td>
<td>msec.</td>
<td>percentage</td>
<td>msec.</td>
</tr>
<tr>
<td>680</td>
<td>PRE-PAUSE</td>
<td>100%</td>
<td>740</td>
<td>PRE-PAUSE</td>
</tr>
<tr>
<td>380</td>
<td>PRE-PHRASE</td>
<td>56%</td>
<td>590</td>
<td>PRE-PHRASE</td>
</tr>
<tr>
<td>350</td>
<td>&quot;</td>
<td>51%</td>
<td>590</td>
<td>&quot;</td>
</tr>
<tr>
<td>350</td>
<td>&quot;</td>
<td>51%</td>
<td>580</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

"THE ELASTIC WORD", JANE GAITENBY, HASESKE LABORATORIES
HAND-OUT #3