

AUGUST 01 1975

Vowel duration in American English

Noriko Umeda



J. Acoust. Soc. Am. 58, 434–445 (1975)

<https://doi.org/10.1121/1.380688>



Articles You May Be Interested In

Consonant recognition and the articulation index

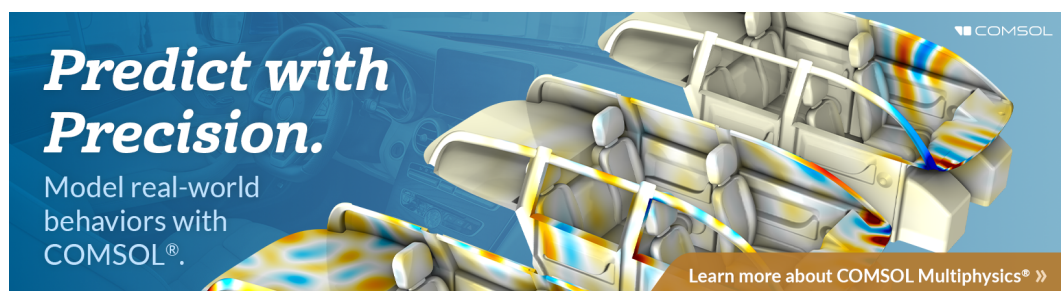
J. Acoust. Soc. Am. (April 2005)

INDSCAL study of the perceptual space of American diphthongs Dale Terbeek

J. Acoust. Soc. Am. (August 2005)

Emotional speech acoustic model for Malay: Iterative versus isolated unit training

J. Acoust. Soc. Am. (October 2013)



Predict with Precision.

Model real-world behaviors with COMSOL®.

Learn more about COMSOL Multiphysics® »

Vowel duration in American English

Noriko Umeda

Bell Laboratories, Murray Hill, New Jersey 07974
(Received 1 January 1975; revised 25 April 1975)

This is a summary report of the vowel duration data that have been accumulated over the past several years. The data corpus analyzed to derive temporal controls of vowels consists mainly of three different readings by three different speakers, each about 10 to 20 min in duration. The rules cover the temporal behavior of vowels under many phonological conditions. The conditions include stressed and unstressed positions, prepausal and nonprepausal positions, word-final and non-word-final conditions, and monosyllabic and polysyllabic words. The influence of following consonants is discussed as well. Included also are conditions other than phonological ones, such as the effect of the prominence of words on their vowels, and the speed of reading. The duration rules derived from the data are intended for use in our speech synthesis-by-rule system from printed text.

Subject Classification: 70.20, 70.70.

INTRODUCTION

A number of papers have reported on vowel duration in English. None of them, however, discussed vowels occurring in continuous, extended spoken materials,¹⁻⁵ and very few of them discussed vowel duration in unstressed positions. The purpose of this paper is to present a large corpus of data for vowel durations in continuous text under as many conditions as possible.

Data shown here were obtained mainly from 10- to 20-min readings by three different speakers. Two speakers, SP and JH, read at their natural reading speed, and the third one, CC, read in a deliberate manner. SP's reading contains the largest number of tokens and shows the most consistency.

It is the nature of language that the probability of occurrences of phonemes and of conditions under which they occur is very uneven. Therefore, even with this large body of data, for a number of conditions there are not enough tokens to provide statistically meaningful estimates.

Durations were measured on sound spectrograms. On the spectrogram, a phoneme boundary is usually determined at a discontinuity in excitation, formant structure, or both. Since this is almost the only way to obtain consistent agreement among investigators, our measurements of vowel duration have followed this method. However, many phoneme boundaries are undefinable on spectrograms. Boundaries between two vowels, and those preceded or followed by /y/, /w/, /r/, and /l/, are typically difficult to determine. These cases were excluded from this report.

There is little agreement on whether the aspiration portion of a voiceless stop should be interpreted as a part of the stop or that of the following vowel. In our measurements, the aspiration is included in the vowel portion for two reasons (for the aspiration time of stops in relation to their closure time, see Ref. 6). First, this interpretation assures consistent measurements among voiceless stops, voiced stops, and nasals: in all of these cases, the transition occurring after the release of vocal tract closure is included in the vowel. The second reason is that the distribution of this total

duration has less variability under any single condition than either aspiration or the "truncated vowel" alone.⁷ At the same time, this interpretation of vocalic duration shows the influence of voiceless stops to be similar to that of other preceding consonants.

I. DURATION OF THE MAIN VOWEL IN THE WORD

A. Positional conditions

Three positional conditions were considered. The *prepausal* condition includes the vowel in the last syllable before a pause. The pause is defined as a termination of a message unit accompanied by a special pitch contour on the last syllable of the unit with or without a physical silence following. A pause without silence is often called "pseudopause" and is found more frequently in normal English reading than the pause with silence.

The *monosyllabic* condition includes the vowel in the monosyllabic word and the stressed vowel in the last syllable of a polysyllabic word, both in nonprepausal positions. In our data of fluent speech, no significant difference is observed in the durational behavior between these two situations.

The third category of the positional conditions is *polysyllabic*. All stressed vowels in polysyllabic words other than those in the above-mentioned categories fall in this category.

Figures 1 and 2 illustrate durations of eleven reasonably frequent vowels by two speakers, SP and JH, respectively, under the three positional conditions. All other conditions, such as following consonants and word prominence, are pooled together here. Vowels that do not occur often enough in the data, such as /u/, /v/, /ɔ/, and /ɔɪ/, are not shown in the figure.

Each data point in these figures indicates the mean duration of the total occurrences of the vowel in that positional condition. This means that all data points are not necessarily statistically and phonologically equivalent. Which conditions occur more frequently depends on the particular vowel, the positions where it may occur, and on individual reading materials.

31 January 2026 16:32:17

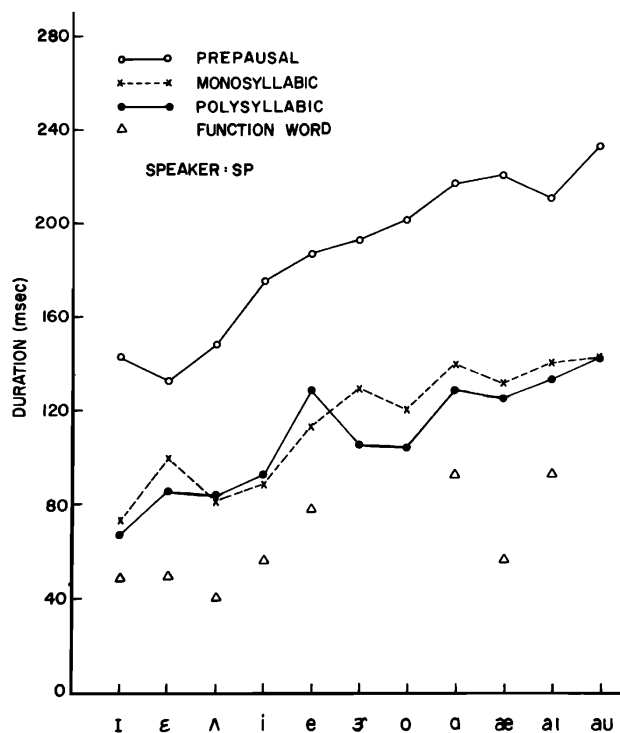


FIG. 1. Average duration of vowels under three positional conditions: prepausal, monosyllabic, and polysyllabic. The speaker is SP.

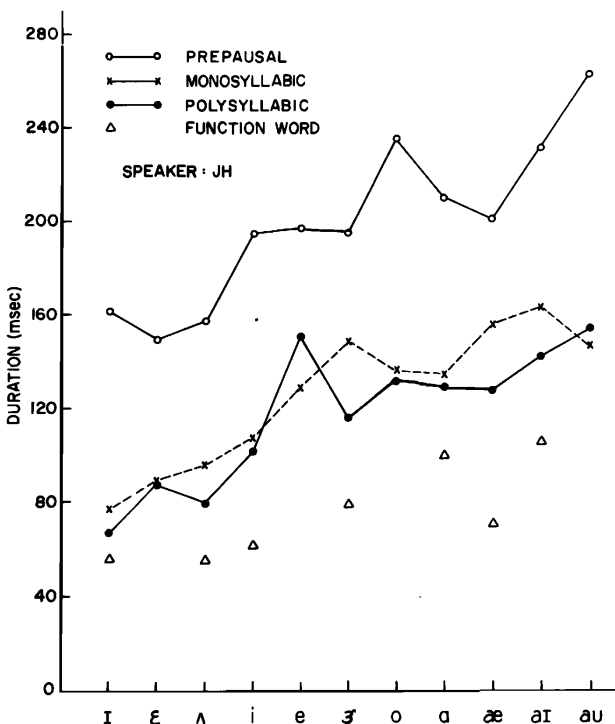


FIG. 2. Average durations of vowels under three positional conditions, by speaker JH.

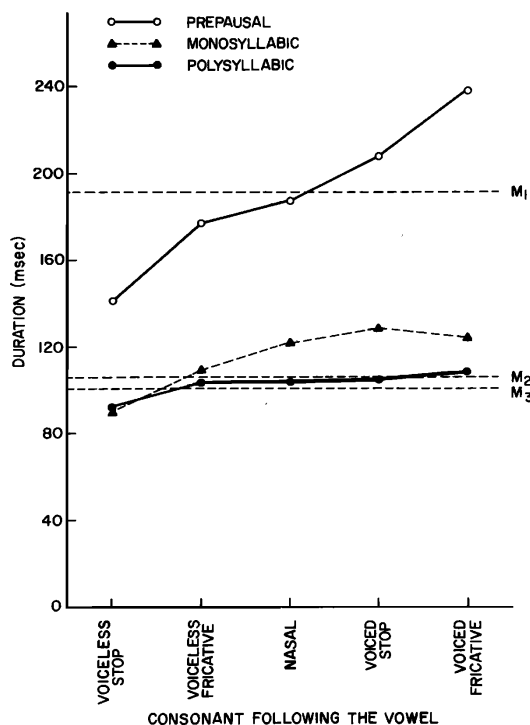


FIG. 3. The effect of following consonants on vowel durations under the three positional conditions. All of the vowels shown in Fig. 1 are pooled together here. The speaker is SP.

Triangles in the figures represent mean values for the duration of vowels in function words. Though these vowels are said to be reduced nearly to schwas in natural situations, their durations display a difference according to their phonemic identities (in detail, see Sec. I-D). Articles, "the" and "a," are excluded from these figures.

B. Consonantal conditions

The well-known fact of the influence of the following consonant on vowel duration holds true with most of our "prepausal" vowels, but the majority of the vowels in running speech do not fall into this situation. This effect on "monosyllabic" and "polysyllabic" vowels is very slight, except when a voiceless stop follows the vowel. Figure 3 shows, for speaker SP, the trend of this effect in terms of positional conditions by pooling all of the vowels together (from Fig. 1). M_1 , M_2 , and M_3 are the mean values of all vowels in *prepausal*, *monosyllabic*, and *polysyllabic* positions, respectively. Cases where an obvious syllable boundary occurs between the stressed vowel and the following consonant within a word (e.g., "peapod") are extremely rare in all of our materials. They are, therefore, excluded from this figure.

Breaking down this data in more detail, we show in Fig. 4 that for each consonantal condition different types of vowels display different durational tendencies. Data on consonantal conditions other than those shown in this figure are lacking for many vowels. The zero point on

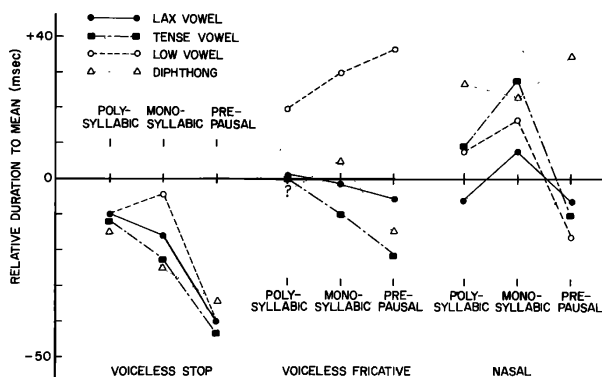


FIG. 4. Detailed differences in consonantal effect among different groups of vowels by speaker SP. Three consonantal conditions that are frequent for all vowels and all conditions are shown.

the abscissa in the figure indicates the mean duration of each vowel under each positional condition. The speaker is SP. The eleven vowels are classified into four categories, according to durational behavior and phonological distribution: lax vowels, /ɪ/, /ɛ/, and /ʌ/, which do not appear in open syllables; tense vowels, /i/, /e/, /o/, and /ɜ/, which can fall in open syllables (the vowel /ɜ/ is not a tense vowel, but its durational behavior is similar to that of tense vowels); low vowels, /æ/ and /a/; and diphthongs, /aɪ/ and /aʊ/.

The figure shows that all vowels behave similarly when voiceless stops follow them. It shows that before voiceless fricatives, low vowels behave completely opposite to other groups of vowels. /æ/ has a deterministic influence on this trend, because this vowel occurs far more frequently than /a/. The fact that /æ/ followed by a voiceless fricative is longer than when followed by any other consonant is a peculiarity of this speaker's dialect (see Figs. 22 and 23).

Before nasals, diphthongs behave differently from other vowels. Vowels followed by nasals, including the ones clustering with voiced and voiceless stops and fricatives, vary more widely on the duration scale (the standard deviation is about 20 msec) than under other conditions (about 10 to 15 msec). The question mark in the figure indicates insufficient tokens for that condition.

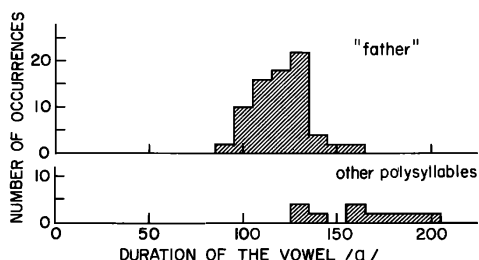


FIG. 5. Illustration of the effect of the reduction of word prominence by repetition in the discourse. The example is the vowel /ɑ/ in "father" compared with the same vowel in other polysyllabic words spoken by JH.

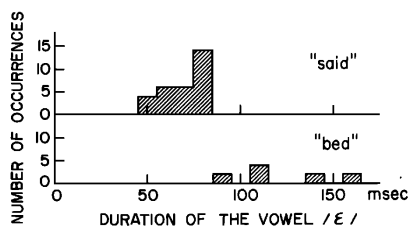


FIG. 6. Difference in vowel duration in terms of the probability of occurrence of the words "said" (1961 out of a million) and "bed" (127 out of a million). The speaker is JH.

Some other vowels do not occur frequently enough before voiceless clusters, voiced stops, or voiced fricatives to produce the same kind of figure.

Some explanation of consonant clusters is necessary. A cluster is named after its first member. Voiced-consonant clusters behave sufficiently like their first member, alone, to be classified into the same category. Nasal clusters do not show a systematic influence according to the voiced-unvoiced distinction of the second member, except for a few conditions (the duration of the nasal itself varies greatly according to this distinction, for example, in /ns/ 47 msec versus /nz/ 73 msec). The influence of voiceless-stop and -fricative clusters on lax vowels and diphthongs is similar to that of single voiceless fricatives. With low and tense vowels, those clusters have to be treated separately, and differently, from each other and from their single first members.

No consistent effect of preceding consonants on vowel duration is seen, except for /h/ preceding the vowel /æ/; in this case the vowel is the shortest of all. The difference in the effect of intervocalic flap /t/ and /d/ on the duration of preceding stressed vowels is not significant in our data.

C. Vowel duration and word prominence

An important factor which affects the duration of vowels in connected speech is "word prominence." This is the attribute related to the information load that the word carries in the message—the measure of unpredictability. This predictability depends on how unusual or important the word is in a particular situation. Words which occur infrequently in English have more specific meanings than frequent ones, and require longer time to recall.^{8,9} Nouns form the largest open class, and so they are hard to guess from context when they are missed in the flow of speech; prepositions are usually very easy to guess from the context. Unpredictable or important words take more exaggerated acoustic attributes than more predictable or less important words. The vowel duration may be included among attributes which are affected by this factor.

Two examples are shown here. Repetition reduces the importance of the word rather drastically, starting from its second occurrence in the discourse. Figure 5 shows that the duration of /ɑ/ in "father," which occurs about 75 times (in the material read by JH), is reduced greatly, compared with the same vowel in other polysyllabic situations.

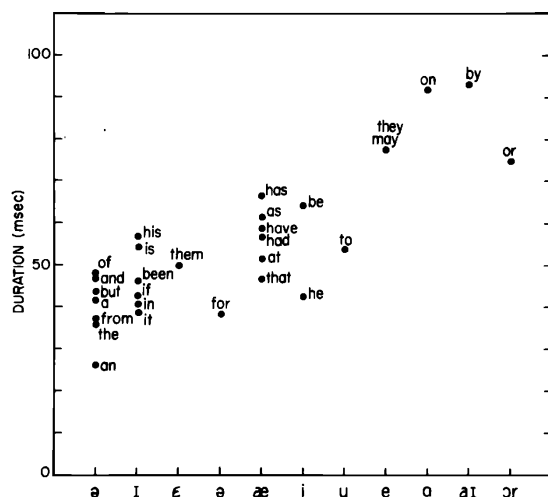


FIG. 7. Average duration of vowels in function words spoken by SP.

The unconditional probability of word occurrence in English is another rough measure of word prominence. The word "said" occurs 1961 times out of a million (53rd in rank in the data by Kucera and Francis¹⁰ and the highest-ranked verb), whereas "bed," which has the same vowel followed by the same consonant, occurs 127 times out of a million. This difference, probably coupled with the grammatical difference, is reflected in the duration distributions of the two words shown in Fig. 6. The extreme example of this kind is seen in function words.

D. Vowel duration in function words

The distinction between function word and content word can not be clearly made in the acoustic realization of speech; rather, the change from extremely functional to very contentful seems to be continuous. But in order to investigate the distinction, it is necessary to draw a dividing line. Therefore, words up to the 50th in rank out of 51 216 words in Kucera-Francis are considered function words in Fig. 7 (personal pronouns up to the 100th rank are included). This criterion includes all short, weak function words, and excludes more specific and polysyllabic function words. It also excludes the most frequent words in open classes; for example, the most frequent verb, "said," appears at 53rd, the most frequent adjective "new" at 64th, and the most frequent noun "time" at 67th.

Figure 7 shows vowel durations of function words in SP's reading. Data are shown for those words of which there were more than ten usable tokens. Many function words begin or end with /w/ or a vowel, and therefore, in many cases, it is impossible to draw boundaries with preceding or following words. The figure clearly illustrates the effect of vowel identity and following consonants. A similar trend is obtained with other speakers. Speaker JH's durations tend to be about 10 msec longer than SP's, and CC's 20 to 30 msec longer.

II. DURATION RULE FOR MAIN VOWELS

Though each vowel has its own pattern of elongation, a similar interaction between consonantal factors and positional factors is observed in all vowels. This suggests the possibility of forming a simple rule for vowel durations. Such a rule for the diphthong /aɪ/ is shown in Figs. 8 and 9. Consonantal factors are plotted on the abscissa, and various position and stress factors are indicated as separate parameters in Fig. 8. The location of these factors is reversed in Fig. 9. Numbers in the figures are counts of occurrence.

The following is the rule for the duration of the diphthong /aɪ/:

$$T = 60 + S(130 + 130 \times C). \quad (1)$$

In this formula, the value of C depends on which consonant follows the vowel. The value of S depends on many factors: the position of the vowel in a word and in a sentence, the word prominence, sentence stress, speech rate, etc. It is a considerable simplification to assume that all these factors work linearly.

Figures 10-23 show data for other vowels similar to that obtained for /aɪ/. The vowels that are not presented here do not have enough occurrences in our data.

Any monophthong in American English, when it appears under conditions which result in lengthening, becomes a diphthong. Or, more strictly, in this situation, the tongue does not stay in a fixed position, but rather moves slowly in a certain direction throughout the vowel. This causes the nonlinearity in our simple rule. In lax vowels, the diphthongization occurs only when they are the longest, namely under a combination of prepausal and voiced-fricative conditions, as observed in /ɪ/ in Figs. 10 and 11 (/ɛ/ and /ʌ/ have no tokens under this condition). Each of the tense vowels shows its own manner of nonlinearity. We see the reason why /aɪ/ follows the rule most regularly: it is a diphthong to begin with.

The strangest vowel is /æ/. The pattern of diphthongization of this vowel (possibly of /a/ and /ɔ/, too) varies greatly from dialect to dialect. And, in fact, the data show a discrepancy between speakers (see Figs. 22 and 23). When a voiceless fricative follows it, this vowel becomes the longest in SP's reading ("educated New York" dialect), but does not in other speakers (JH is from Ohio, and CC from the southern U.S.). Voiceless-fricative clusters (as in "past") elongate the vowel far more in SP's reading.

Table I shows values of three constants, T_0 , K_1 , and K_2 , in Eq. 2 for eight vowels:

$$T = T_0 + S(K_1 + K_2 \times C). \quad (2)$$

Table II shows the C factor for each vowel under each consonantal condition. Table III lists the S factors by vowel and positional condition.

Table I shows that /æ/ has the steepest slope for both C and S factors. This is explained by a great difference between vowel durations in polysyllabic content words

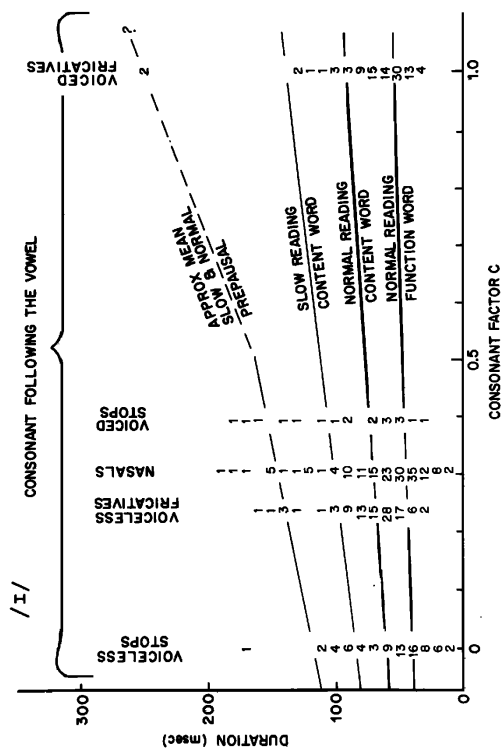


FIG. 10. Duration of /ɪ/ as a function of the constant that follows it.

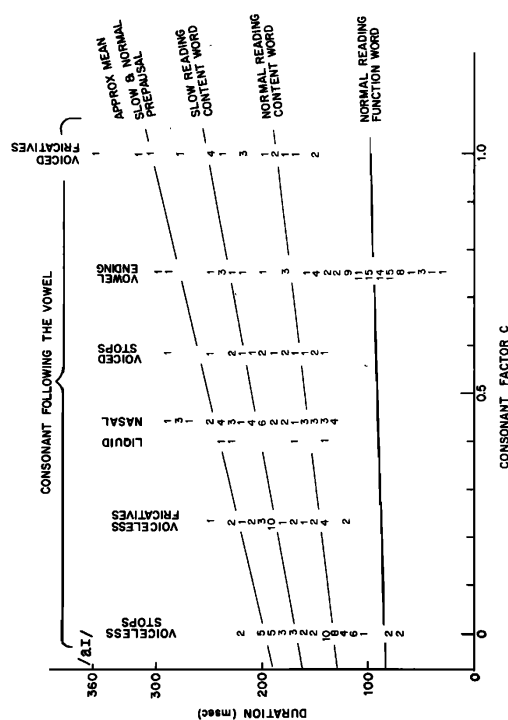


FIG. 8. Duration of /aɪ/ as a function of the constant that follows it. Parameters in the figure are approximate means for stress-situation factor in Fig. 9. SP and JH read essays at their normal speeds, and CC very slowly read three stories consisting of monosyllabic words.

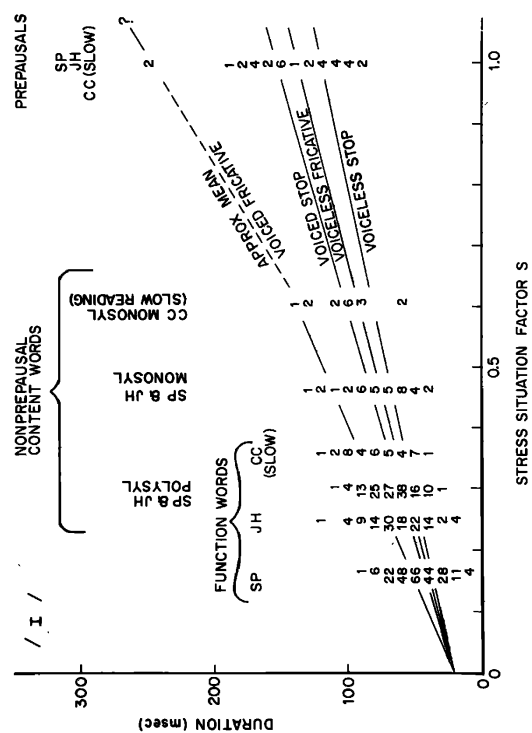


FIG. 11. Duration of /ɪ/ as a function of stress and situation.

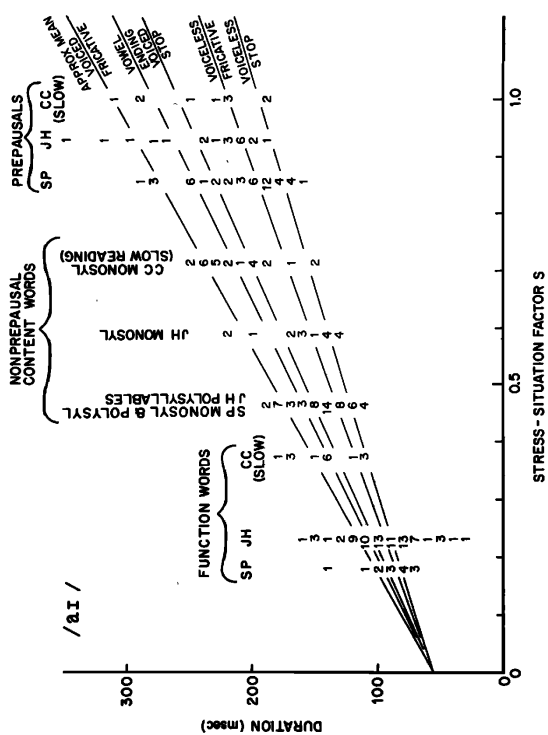


FIG. 9. Duration of /aɪ/ as a function of stress and position in the word and the sentence. Parameters in the figure are approximate means of the consonantal factors in Fig. 8.

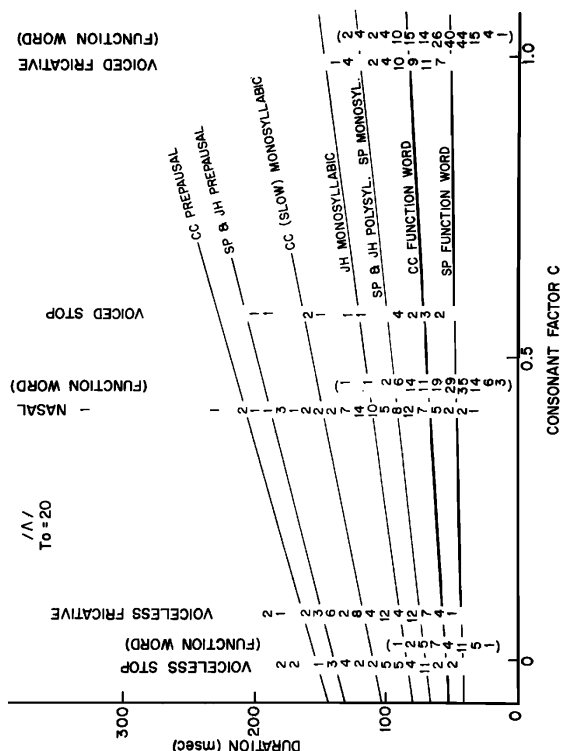


FIG. 14. Duration of /a/ as a function of the following consonant.

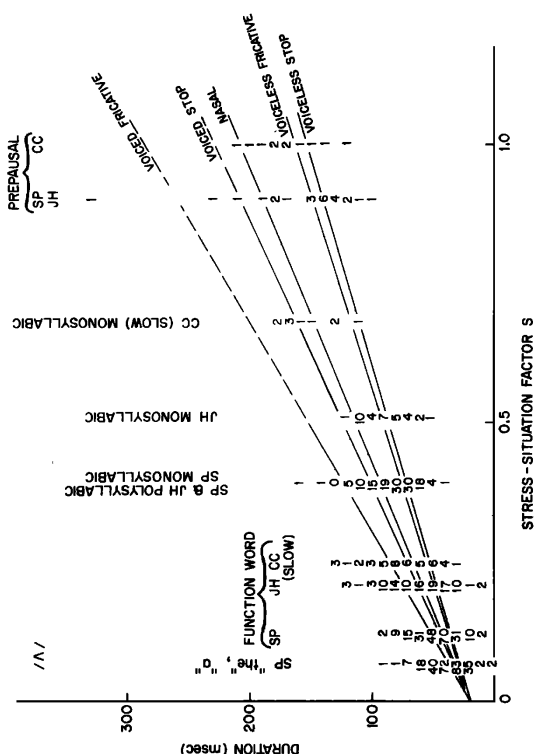


FIG. 15. Duration of /a/ as a function of stress and situation.

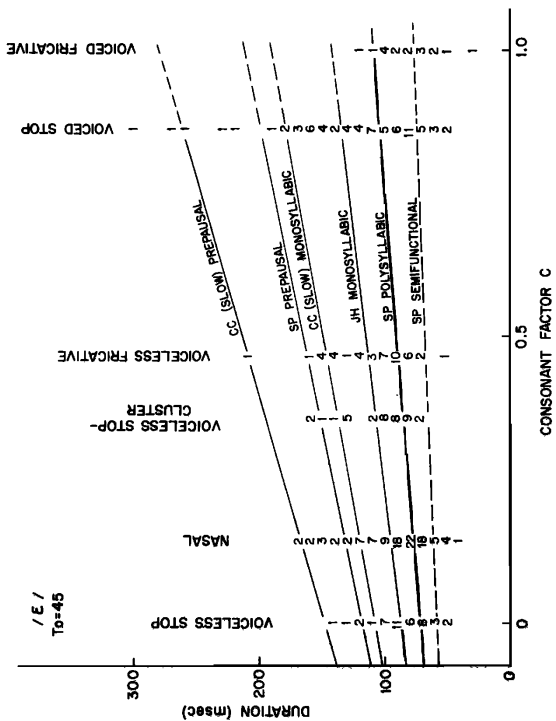


FIG. 12. Duration of /ε/ as a function of the consonant that follows it.

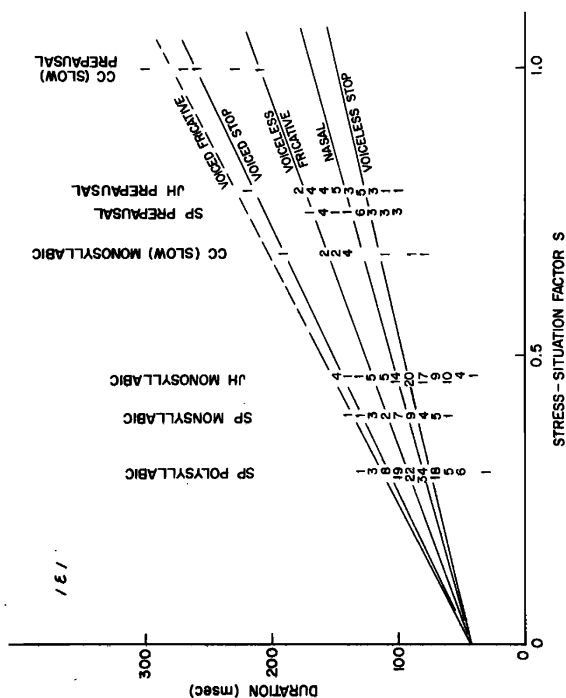


FIG. 13. Duration of /ε/ as a function of stress and situation.

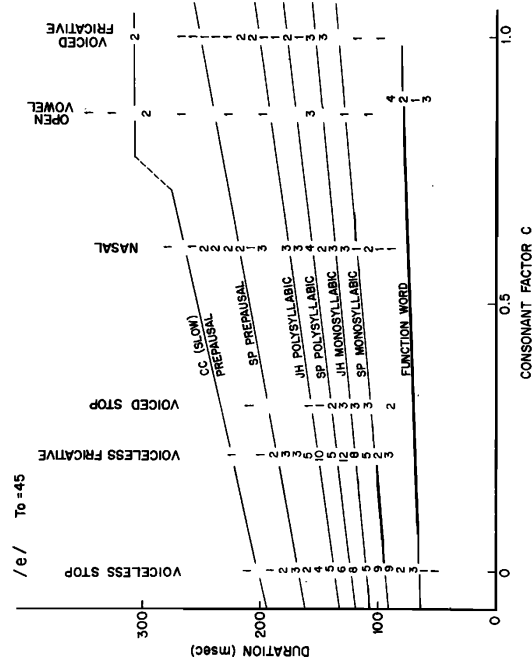


FIG. 16. Duration of /i/ as a function of the following consonant.

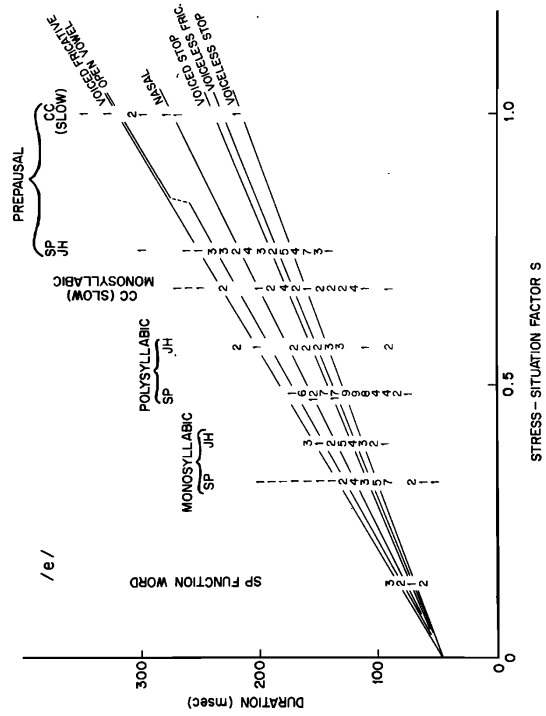


FIG. 17. Duration of /i/ as a function of stress and situation.

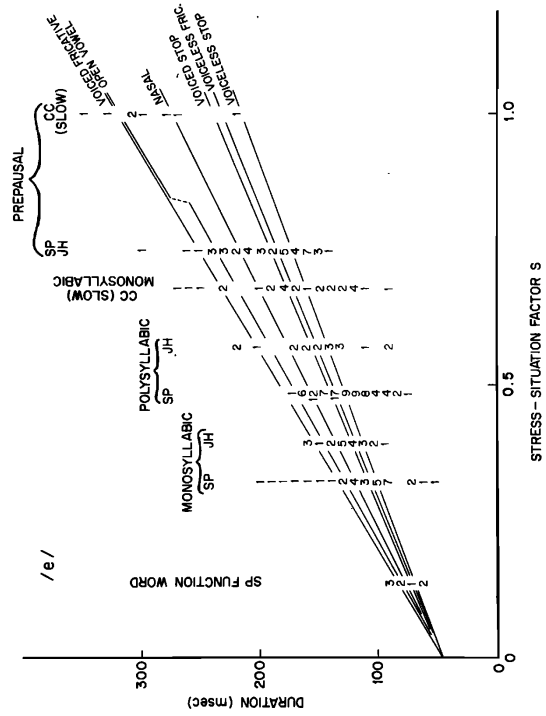


FIG. 18. Duration of /e/ as a function of the following consonant.

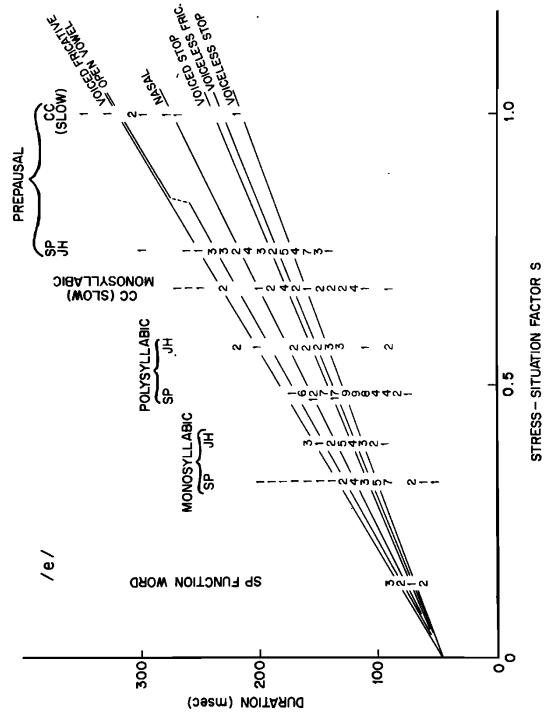


FIG. 19. Duration of /e/ as a function of stress and situation.

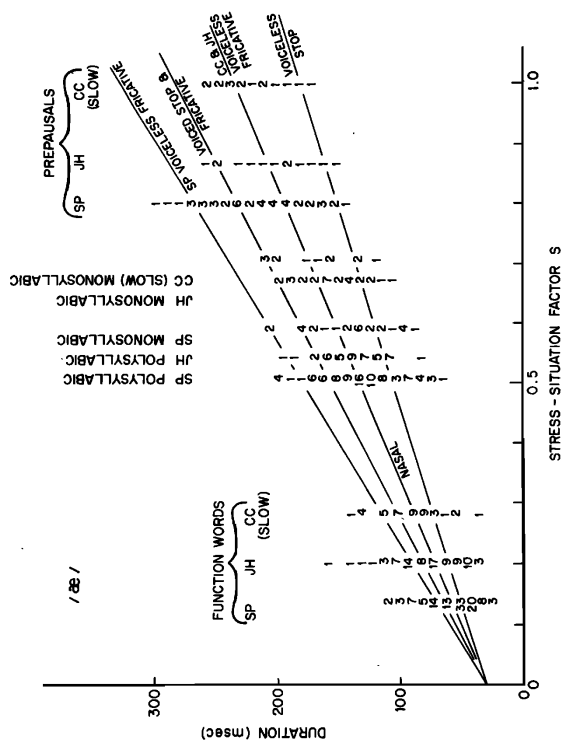


FIG. 22. Duration of /æ/ as a function of the following consonant.

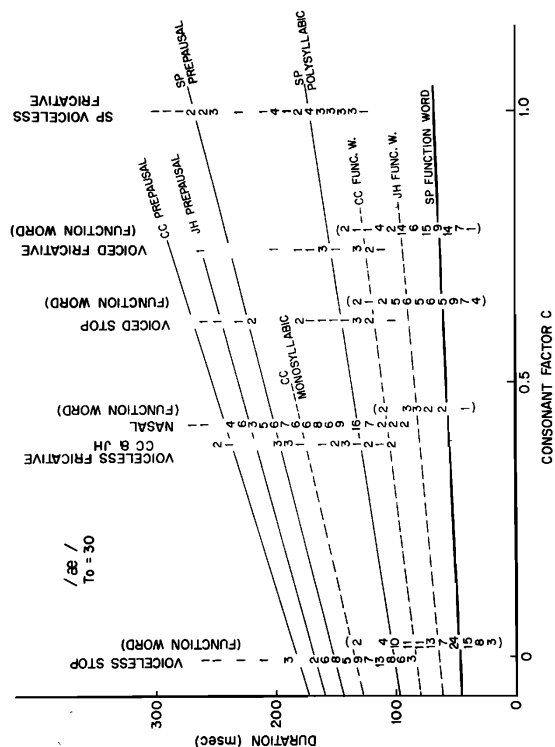


FIG. 23. Duration of /æ/ as a function of stress and situation.

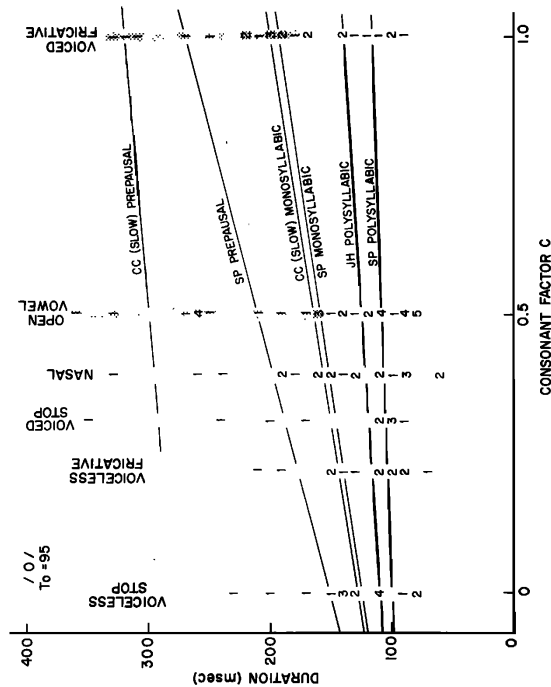


FIG. 20. Duration of /o/ as a function of the following consonant.

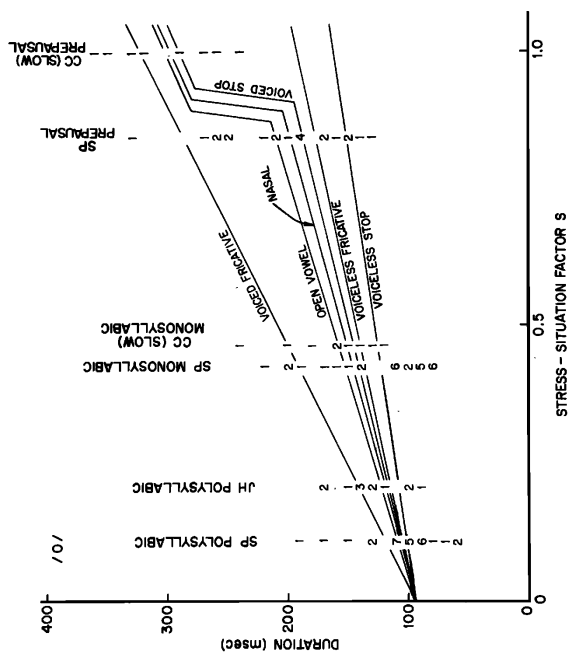


FIG. 21. Duration of /o/ as a function of stress and situation.

TABLE I. Values of three constants T_0 , K_1 , and K_2 .

	T_0	K_1	K_2
i	20	100	100
e	45	130	105
æ	25	135	130
i	60	110	130
ε	45	105	170
Δ	95	160	70
æ	30	145	155
aɪ	60	130	130

and those in function words (see Table III and Figs. 1 and 2). The vowel /i/ shows the most gradual slope for both factors. This vowel is the least sensitive to almost any elongation factor except for the voiced-fricative condition in the prepausal situation (see Figs. 10 and 11).

Tables II and III show that the three lax vowels /ɪ/, /ε/, and /Δ/ behave alike except for the prepausal ones, where a difference between speakers is observed. In the diphthong /aɪ/, all conditions are distributed evenly on the scale of any factor. Peculiarities seen in Tables II and III are that for the voiced-stop condition of /ε/ and /e/ the duration is very short, and that /e/ in the

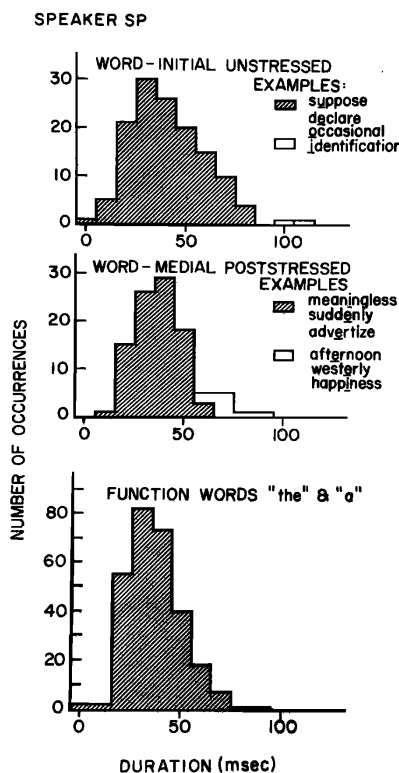


FIG. 24. Duration distribution of unstressed vowels in non-word-final positions, compared with the vowel in the most reduced function words "the" and "a." Blank areas in the figure are tokens under conditions that require longer duration. The speaker is SP.

SPEAKER JH

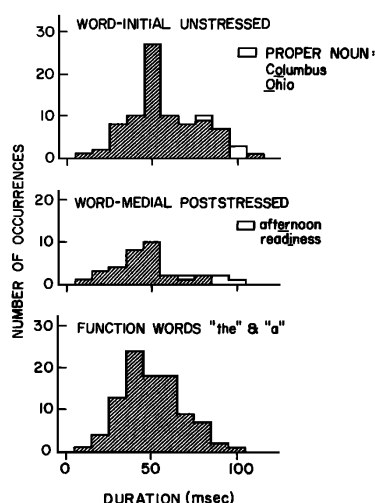


FIG. 25. Data distribution for the same conditions as those in Fig. 24. The speaker is JH.

polysyllabic condition is longer than in the monosyllabic condition.

III. VOWEL DURATION IN UNSTRESSED SYLLABLES OF POLYSYLLABIC WORDS

Vowels without primary stress in polysyllabic words are (1) schwas or reduced vowels in unstressed syllables, (2) vowels which are unstressed, but retain their vowel identities ("realize," "identification," "ancestor," etc.), and (3) vowels with secondary stress.

Classes (2) and (3) are extremely infrequent, and therefore inappropriate for a thorough study with our data. The first class, unstressed vowels, is subclassified into word-final and non-word-final conditions. Vowels in word-final positions are longer than those in other positions, and clearly show the effect of the following consonant.

A. Unstressed vowels in non-word-final positions

In Fig. 24, unstressed vowels in three situations are compared from top to bottom: vowels in word-

TABLE II. C-factor value for each vowel.

	Voiceless stops	Voiceless fricatives	Nasals	Voiced stops	Vowel ending	Voiced fricatives
i	0.	0.23	0.30	0.39	...	1.00
ε	0.	0.47	0.14	0.36	...	1.00
Δ	0.	0.08	0.43	0.57	...	1.00
i	0.	0.11	0.44	0.68	0.97	1.00
e	0.	0.21	0.60	0.30	0.85	1.00
o	0.	0.21	0.40	0.61	...	1.00
æ	0.	0.40	0.43	0.61	...	0.73
aɪ	0.	0.23	0.44	0.59	0.75	1.00

^aVoiced-stop clusters.

^bSP.

TABLE III. S-factor value for each vowel.

	Function words			Polysy		Monosyllabic			Prpausal		
	SP	JH	CC	SP	JH	SP	JH	CC	SP	JH	CC
¹	0.16	0.24	0.36	0.30	0.30	0.47	0.47	0.60	1.00	1.00	1.00
^e	...	...	...	0.30	...	0.40	0.45	0.68	0.75	0.79	1.00
^a	0.07 ^a 0.11 ^b	0.20	0.25	0.39	0.39	0.39	0.50	0.68	0.90	0.90	1.00
ⁱ	0.00	0.00	0.15	0.18	0.18	0.18	0.30	0.40	0.53	0.63	1.00
^e	0.13	...	...	0.48	0.57	0.32	0.40	0.68	0.75	0.75	1.00
^o	...	...	...	0.10	0.20	0.43	...	0.47	0.85	...	1.00
^æ	0.13	0.20	0.28	0.50	0.55	0.59	0.68	0.71	0.80	0.87	1.00
^{aɪ}	0.18	0.23	0.37	0.46	0.46	0.46	0.60	0.71	0.85	0.93	1.00

^a"The" and "a."^bOther function words.

initial syllables, vowels in a word-medial syllable which immediately follows a stressed syllable, and vowels in the shortest function words "the" and "a." The speaker is SP. The same situations are illustrated for JH in Fig. 25.

The two figures indicate that very unstressed vowels have similar durations if they are not word terminal. Articles are always pronounced as if they were a part of their following content words, and in this sense their vowels are like those in word-medial position.

In both speakers, vowels in these articles are the shortest, and word-medial, poststressed ones are slightly shorter than word-initial ones. The data also indicate that word-initial vowels have a wider range than word-medial ones. The vowels in word-medial position immediately after a stressed syllable are usually the weakest in the rhythm structure of English; their duration is little affected by their phonemic contexts, stress, or boundaries. On the other hand, vowels in the word-initial position are affected by various factors such as word prominence and grammatical and semantic boundaries. A wider variation in the JH durations occurs because he began his 15-min reading session in a careful, slow manner and ended it in a somewhat sloppy manner. SP controlled his speed and manner consistently.

In the word-initial section of Figs. 24 and 25, blank tokens are those that retain their vowel identities. They happen to have the same duration as the mean duration of the vowel in "by" (see Fig. 7). The blank areas in word-medial, poststressed vowels indicate vowels that fall in an open syllable followed by an initial-consonant allophone in the suffix.⁶ These vowels show the same tendency as the word-final ones that will be discussed in the following section.

B. Unstressed vowels in word-final syllables

Unstressed vowels occurring in the final syllable of a word show a considerable difference in the effect of the following consonant, depending on whether or not the syllable is in the prepausal position. Figure 26 il-

lustrates this phenomenon. Vowels of nonprepausal suffixes are plotted on the abscissa and those of prepausal suffixes on the ordinate. Solid circles are taken from SP's reading and blank circles are the averages of three other speakers reading at their normal speed. The definition of "prepausal" is the same as that in main vowels.

The nonprepausal category, where no conspicuous pitch change is observed at the end of the word, also contains a factor of elongation. Such a factor is due to a "semantic separation," where two important or infre-

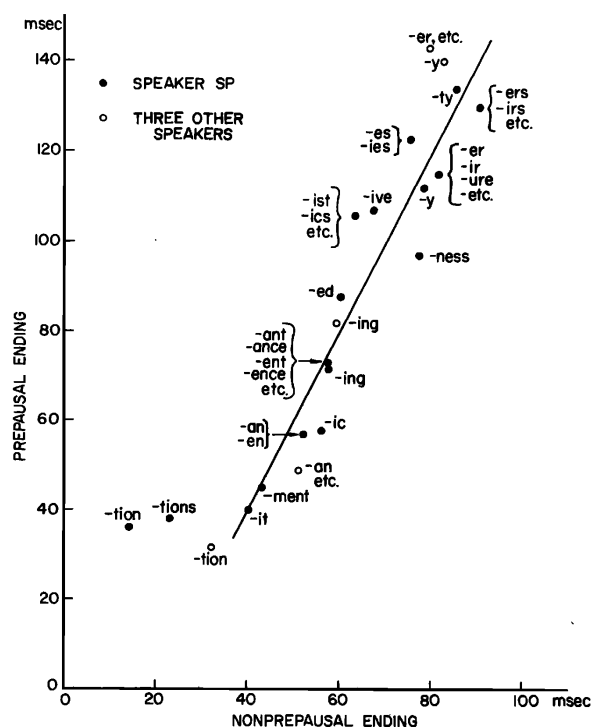


FIG. 26. Duration of vowels in suffixes. Duration of prepausal vowels are plotted on the ordinate and nonprepausal ones on the abscissa.

quent words form a phrase. In this case, the vowel of the last syllable of the first word is elongated—for example, in “vulgar dream,” /ə/ is the longest (110 msec) among nonprepausal /ə/'s.

In Fig. 26, the line showing a 2:1 relationship between *prepausal* and *nonprepausal* vowels is only intended to give a rough picture of the difference between these situations. The lengthening order among consonants is as follows: /n/ < voiceless stops < nasal other than /n/ and nasal clusters < /s/ and voiceless clusters < open vowels < voiced fricatives. Suffixes “-tion” and “-tions” fall off from the general trend. This is because, in many nonprepausal situations, no vowel portion of these suffixes was observed on the spectrogram (there was no such case in prepausal ones). These cases were counted as duration “zero” and averaged with the cases where some vowel portion was observed.

IV. THE VARIABILITY OF THE DATA

There is, of course, a considerable amount of variability in the data. However, with more than five hundred conditions it is impossible to discuss their variability in great detail. Therefore, we present a summarized discussion for a rough estimate of the variability of three situations: main (stressed) vowels, unstressed vowels in polysyllabic words, and vowels in function words. The variability of data presented here is based on SP's reading. For other speakers, JH and CC, standard deviations are 2–4 msec greater than those of SP's data.

Among stressed vowels, front vowels show less variability than low or central vowels (back vowel are very infrequent). Both under *monosyllabic* and *polysyllabic* conditions (i.e., excluding prepausal vowels), standard deviations for high vowels are about 13 to 15 msec, depending on vowel identities and the kind of consonant following them. For low and central vowels, the average standard deviation is about 20 msec. Under *prepausal* conditions, the standard deviation for all vowels ranges from 22 to 29 msec.

Unstressed vowels in the word-initial position show 17 msec as their standard deviation; and unstressed vowels in word-medial, poststressed positions 11 msec.

In word endings, vowels in short and medium length (see Fig. 26) demonstrate 13 msec as the average standard deviation with the range of 7 to 17 msec, for both prepausal and nonprepausal conditions. Vowels of greater duration (-er, -ers, -y, -ies, etc.) have greater standard deviations: 20 msec for nonprepausal vowels and 25 msec for prepausal ones.

Vowels in function words demonstrate a similar range of variability to vowels in other situations. The /ə/ group (the, a, but, etc.) and the /ɪ/ group (it, is, his, been, etc.) show standard deviations of 7 to 17 msec. The /æ/ group (have, has, as, at, etc.) shows greater standard deviations, ranging from 14 to 20 msec. The majority of all function words have standard deviations in the range of 10 to 15 msec.

V. SUMMARY

This paper contains a description of vowel duration data from natural readings by several speakers. The data are classified mainly into various detailed phonological conditions. However, there are many more factors in speech that determine vowel duration than we have discussed in this paper. For instance, the distinction between content words and function words in our rule is too coarse a quantization on the scale of word prominence. Another factor is that various degrees of continuation at the grammatical boundary are observed; this affects the duration of prepausal vowels accordingly. Such factors, which to a considerable degree result from subjective choices reflecting the speaker's intention, are difficult to objectively sort out from the speech wave. They may cause a wide distribution in our data.

Some phonological factors that may yield different influences on vowel duration have been pooled together because of their insufficient occurrences in our materials. A considerable number of vowels and conditions are omitted from this paper; for example, vowels with secondary stress, and unstressed vowels in situations other than those presented here.

In spite of such disadvantages, which are inevitable in the study of natural speech materials, the rules derived from this detailed analysis of phonological conditions show an encouraging consistency. (The comparison of these results with those obtained from carrier sentences has been discussed elsewhere.¹²) We believe that these rules are a reflection of performance of the speaker's subconscious control of temporal factors in speech, albeit a considerably simplified version. The rules will become more powerful for speech synthesis and recognition when the influences of higher-level prosodic features, such as stress and boundary, and of the syntactic and semantic structure of the message are taken into account.

These rules will be incorporated in the scheme of our speech synthesis-by-rule system from printed text.

ACKNOWLEDGMENT

I express my deepest gratitude to Cecil Coker for many useful and stimulating discussions and suggestions on this topic. I also appreciate my colleagues' help in making thousands of spectrograms.

¹A. S. House and G. Fairbanks, “The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels,” *J. Acoust. Soc. Am.* 25, 105–113 (1953).

²G. E. Peterson and I. Lehiste, “Duration of Syllable Nuclei in English,” *J. Acoust. Soc. Am.* 32, 693–703 (1960).

³A. S. House, “On Vowel Duration in English,” *J. Acoust. Soc. Am.* 33, 1174–1178 (1961).

⁴D. J. Sharf, “Vowel Duration in Whispered and in Normal Speech,” *Lang. Speech* 7, 89–97 (1964).

⁵T. P. Barnwell, “An Algorithm for Segment Duration in a Reading Machine Context,” *Tech. Rep. 479, MIT Res. Lab. Electron.*, Cambridge, MA (1971).

⁶N. Umeda and C. H. Coker, “Allophonic Variation in American English,” *J. Phon.* 2, 1–5 (1974).

⁷In the cases where the total duration of aspiration and voicing

is shorter than 80 msec, a compensatory phenomenon is noticed which ensures minimal duration of voicing: that is, the aspiration is shortened and voicing part is lengthened, each about 20 msec. Typical examples are the lax vowel /ɪ/ followed by a voiceless consonant (e.g., "kitchen") and unstressed vowels in word-initial positions (e.g., "con-").

⁸W. A. Bousfield and B. M. Cohen, "The Occurrence of Clustering in the Recall of Randomly Arranged Words of Different Frequencies of Usage," *J. Gen. Psychol.* 52, 83-95 (1955).

⁹J. F. Hall, "Learning as a Function of Word Frequency," *Am. J. Psychol.* 67, 138-160 (1955).

¹⁰H. Kucera and W. N. Francis, *Computational Analysis of Present-Day American English* (Brown U. P., Providence, RI, 1967).

¹¹This formula was suggested by C. H. Coker.

¹²M. S. Harris and N. Umeda, "Effect of Speaking Mode on Temporal Factors in Speech: Vowel Duration," *J. Acoust. Soc. Am.* 56, 1016-1018 (1974).