Motor Programs in Rapid Speech: Additional Evidence

Saul Sternberg

Charles E. Wright

Ronald L. Knoll

Bell Laboratories, Murray Hill, New Jersey

Stephen Monsell

University of Chicago

Reprinted from
R. A. Cole (ed.)
Perception and production of fluent speech
(Fourteenth Annual Carnegie Symposium on Cognition)
Pp. 507-534

For reprints: S. Sternberg
2D-446, Bell Laboratories
Murray Hill, NJ 07974

16 Motor Programs in Rapid Speech: Additional Evidence

Saul Sternberg

Charles E. Wright

Ronald L. Knoll

Bell Laboratories, Murray Hill, New Jersey

Stephen Monsell

University of Chicago

I. INTRODUCTION

In a previous article we reported a series of studies of the timing of rapid prespecified action sequences in speech and typewriting (Sternberg, Monsell, Knoll, & Wright, 1978; reprinted as Chap. 15 in the present volume). Under pressure to finish quickly, subjects responded to a signal by typing lists of letters or speaking lists of words that had been specified before the signal. Increments in sequence length increased the time to initiate the first response element (latency effect) and also increased the average time from one element to the next (duration effect). Our experiments permitted us to reject several interesting explanations of the latency effect, and the generality, robustness, and quantitative simplicity of the two effects encouraged us to develop a model for the latency and duration of rapid movement sequences that explained both phenomena in terms of retrieval of ''subprograms'' from a ''motor program'' for the whole sequence that had been compiled in advance. Our results also permitted us tentatively to identify a response *unit* in speech—the *stress group*—in terms of which the observed effects can be most simply described.

In this chapter we provide informal accounts of several studies of speech production that generalize and extend the earlier work and permit us to elaborate our model, defend it against a new contender, and test it further. In providing these accounts we depend heavily on Chapter 15, especially its Sections II, III, and VII, and refer to it extensively (as "LD," usually followed by a section number).

We first demonstrate the generality of the latency and duration phenomena over longer lists than previously reported, different sets of words, and conditions that produce high time uncertainty for the reaction signal (Sec. II). We then summarize findings about the effects of time uncertainty on parameters of latency and duration functions and thereby provide evidence against one class of explanations of the latency effect (Sec. III).

In Section IV we examine the effects of utterance length on the entire latency distribution, rather than just its mean, to provide assurance that the latency effect is not a consequence of a small proportion of long response delays. We then (Sec. V) report measurements of the effects of utterance length on initial values of two acoustic characteristics—fundamental frequency (F_0) and amplitude. The F_0 measurements permit us to reject an alternative explanation of the latency effect—based on peripheral biomechanics rather than central information processing—that was advanced by K. N. Stevens.

One pervasive property of "natural" speech is the decline in F_0 from beginning to end of an utterance; the effect of utterance length on the rate of declination has been taken to indicate advance planning of entire utterances. In Section VI we ask whether the unstructured rapidly spoken lists in our experiments reveal a similar declination effect.

In Section VII we report measurements that permit us to segment rapid utterances into their component words, and thereby to specify how the time interval from one word to the next depends on the serial position of the word pair within an utterance, as well as on utterance length. Although differing in detail, the resulting serial-position functions are similar in important respects to those we obtained in typewriting, and provide further evidence for the dependence of local features throughout the utterance on a representation ("motor program") of the entire utterance.

In Section VIII we report a first attempt to localize, *within* the time interval from the start of one response unit to the start of the next, the effects of utterance length and serial position; these measurements then permit relatively clean tests of two implications drawn from our model.

We turn finally (Sec. IX) to the role of lexical memory in the control of rapid speech, which we investigate by searching for qualitative and quantitative alterations in the latency and duration functions that arise when speakers are asked to produce utterances containing nonwords rather than words.[1]

II. RECITING OF LETTER AND DIGIT LISTS FOLLOWING A RANDOMLY VARIED FOREPERIOD

In the experiments reported in LD we attempted to minimize subjects' time uncertainty by using warning signals and a fixed foreperiod before the reaction signal; we tried to deal with the problem of "anticipation" of the reaction signal by introducing occasional "catch" trials, on which the reaction signal was omitted. In an earlier experiment we had incorporated neither warnings nor catch trials, had varied the foreperiod randomly over 15 equally spaced values ranging from 2.6 to 5.4 sec, and had used lists of letters or digits containing one to six different randomly ordered items; otherwise the procedure was as described in LD, II. Subjects were the same four who served in the speech experiments reported in LD, but before they had had such extensive practice in the task. We include the results here to demonstrate the generality of the latency and duration phenomena to longer lists, different sets of words, and conditions of high time uncertainty.[2] (In Section III we explicitly compare these findings to corresponding data collected under our more typical conditions of low time uncertainty.)

Mean results from the second and third day of the experiment are represented by the open circles in Fig. 16.1. Data shown are for correct responses only; error rates were negligible for $n \leq 5$, but for $n = 6$ the error rate was 9.6%.

Mean latency (Fig. 16.1a) again increased approximately linearly with list length; the slope is 7.3 ± 0.7 msec/word. The linear function accounts for 97.4% of the variance, and no higher order regression coefficients are significant.

Mean duration (Fig. 16.1b) is significantly nonlinear, but is well described by the fitted quadratic function, with 99.96% of the variance accounted for and no significant regression coefficients of order higher than

---

[1]The work described in Sections II–IV of the present chapter will be reported in detail elsewhere by Monsell and Sternberg; the work described in Sections V–IX will be reported in detail elsewhere by Sternberg, Knoll, and Wright.

[2]Let the range of list lengths be $1 \leq n \leq m$. Because the fitting of linear (quadratic) functions to speech latency (duration) data requires estimation of two (three) parameters, only $m - 2$ $(m - 3)$ degrees of freedom remain for evaluating goodness of fit. Increasing m from $m = 4$ and 5 (as in LD) to $m = 6$ therefore substantially increases the sensitivity of such evaluations.

quadratic. Given the larger range of list lengths relative to those of the experiments in LD, this finding strengthens our claim that the duration function is quadratic.

Again, the slope of the latency function (7.3 msec/word) and the quadratic coefficient of the duration function (7.7 msec/word^2) are close in value. However, both parameters are smaller than the corresponding values in other, later speech experiments with the same subjects (see LD, Table 15.1). Another notable difference is the one-intercept of the latency function; its value of 284.9 msec is about 20 msec greater than in the later experiments (see LD, Table 15.1).[3] We show in the next section that only the last of these differences between experiments can be ascribed to time uncertainty.

III. EFFECTS OF TIME UNCERTAINTY ON LATENCY AND DURATION FUNCTIONS

To evaluate performance when time uncertainty was decreased, we also studied the same subjects with the same kinds and lengths of lists as described above, but in a procedure with a fixed foreperiod, countdown signals, and catch trials, like that of our later experiments. (See LD, II.) The procedures with fixed and varied foreperiod were run on different days in a balanced order, but data were collected in the former on only one day. Furthermore, with fixed foreperiods, error rates for lists of lengths 5 and 6 were unpleasantly high (8.0% and 14.8%, respectively), making the latency and duration data somewhat suspect. Data for correct responses are represented by the filled circles in Fig. 16.1.

The slope of the fitted linear latency function is 6.2 ± 0.8 msec/word; the difference of 1.1 ± 0.7 msec/word between slopes for fixed and varied foreperiods in this experiment is neither large nor statistically significant. The difference in latency-function slopes between this and later experiments is not, therefore, a consequence of time uncertainty. (Its source remains an open question.) As one might expect, however, introducing time uncertainty produced a large and statistically significant increase in mean latency (29.0 ± 3.4 msec, averaged over list lengths). Time uncertainty and list length therefore appear to have additive effects on mean latency. Evidence for such additivity supports models in which the two factors influence different processing "stages" (Sternberg, 1969). Additivity also argues against hypotheses about the latency effect such as

[3]The one-intercept, a convenient measure of the height of the latency function, reflects the time to start producing a single word, and is estimated by the parameter η in the latency functions of LD, Table 15.1.

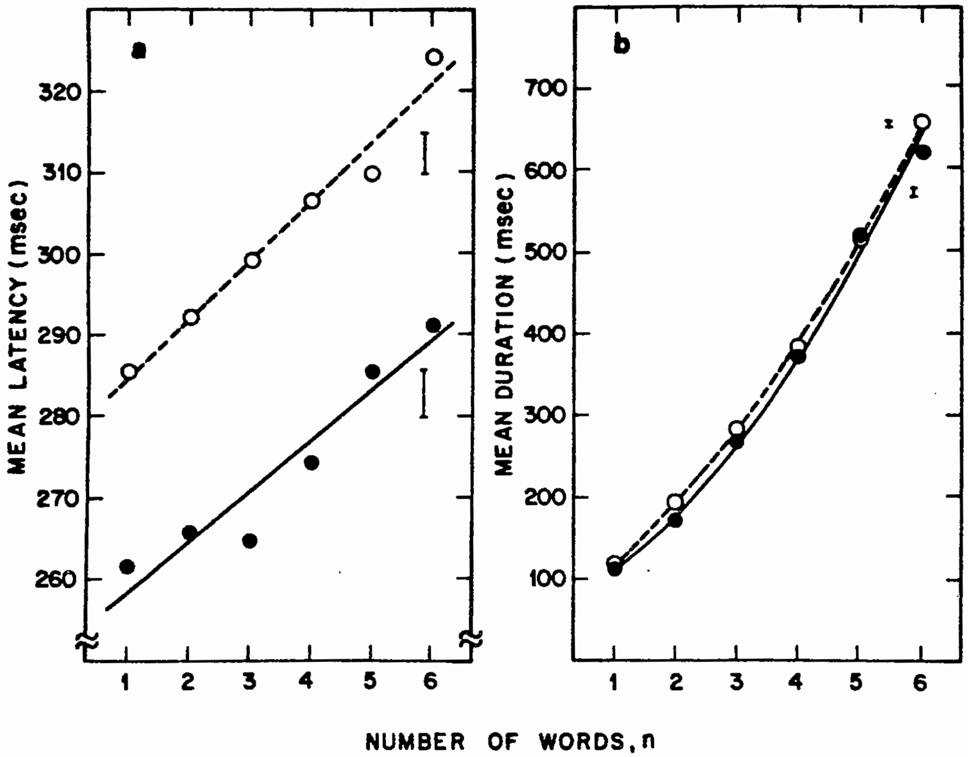

FIG. 16.1. Fixed- and random-foreperiod experiments. Open circles and broken lines show results from the random-foreperiod experiment averaged over four subjects, two vocabularies (letters, digits), 15 foreperiods, and days 2 and 3; about 240 observations per point. Filled circles and unbroken lines show results from the fixed-foreperiod experiment averaged over four subjects and two vocabularies; about 80 observations per point. a. Mean latencies, estimates of standard error ($\pm SE$), and fitted linear functions (random foreperiod, $284.9 + 7.3(n - 1)$ msec; fixed foreperiod, $258.4 + 6.2(n - 1)$ msec). b. Mean durations, estimates of $\pm SE$, and fitted quadratic functions [random foreperiod, $117.1 + 67.9(n - 1) + 7.7(n - 1)^2$; fixed foreperiod, $110.9 + 53.7(n - 1) + 10.6(n - 1)^2$ msec]. See LD, Table 15.1, for fitting procedure.

"time-sharing" (LD, III), in which the effect of list length results from the subject's being unprepared for the reaction signal, and therefore being in a "rehearsal" state rather than a "readiness" state when the signal is presented. (If the effect of list length on latency resulted from such lack of preparation, we would expect any factor that increased time uncertainty also to increase the length effect.)

Mean duration in the fixed-foreperiod procedure (Fig. 16.1b) is again significantly nonlinear, and with no significant regression coefficients of order higher than quadratic. The fitted quadratic function accounts for 99.94% of the variance; deviations for $n \geqslant 5$, which are larger than usual, may be related to the high error rates. Because of these deviations, our estimate of the quadratic coefficient—10.6 ± 1.6 msec/word^2—is more dependent than usual on the choice of estimation method.[4] Hence, although its value is closer than the varied-foreperiod experiment to those found in other experiments, the difference between coefficients cannot be taken seriously. The mean duration difference (averaged over list lengths) of 16.7 ± 12.5 msec is not statistically significant.

IV. EFFECTS OF LIST LENGTH ON THE DISTRIBUTION OF UTTERANCE LATENCIES

Especially because the effect of list length on mean utterance latency is small, with estimated slope values never exceeding about 15 msec/word, we felt it was important to examine its effect on other characteristics of the latency distribution in addition to its mean. One question this would enable us to answer is whether the small effect on the mean is a consequence of mixing a large effect on a few trials with no effect on most trials. (For example, if subjects were occasionally delayed by having to inhale after the reaction signal, despite our instructions to breathe either between trials or before the reaction signal, this would add a large increment to the latency.[5] If for longer lists they either breathed more deeply and hence for more time, or breathed with higher probability, this could have caused the increase in mean latency with list length.)

We used latency data from the experiment with lists of weekdays in normal, random, and repeating sequence (LD, II). For each subject and list length, we pooled the latencies of all correct responses over the three types of sequence, seven starting words, and three replications, to provide about 60 observations per distribution. We then obtained the maximum, mean, and minimum of each distribution, as well as estimates of a set of quantiles, $\{t_p\}$, where the proportion, p, ranged from $p = .05$ to $p = .95$. (If L is a random variable representing the latency, then the p-quantile of its distribution is defined as the latency value, t_p, for which

[4]The procedure for estimating the fitted duration functions whose equations are given in the caption of Fig. 16.1b (LD, Note to Table 15.1) uses the empirical *rate* function to estimate β and γ. When instead, the empirical *duration* functions were fitted directly, parameter estimates from the random-foreperiod data remained approximately invariant $[119.3 + 63.6(n - 1) + 8.6(n - 1)^2]$, but estimates from the fixed-foreperiod data were altered $[103.9 + 62.5(n - 1) + 7.2(n - 1)^2]$. (The two estimation procedures can be regarded as differing in the relative weights placed on data from different list lengths.)

[5]We believe (but have not determined experimentally) that the time increment associated with even a shallow inhalation would be large relative to the spread of the latency distributions ($SD \simeq 30$ msec). See Conrad & Schönle (1979) for supporting evidence.

$\Pr\{L \leq t_p\} = p$.) Means of these statistics over subjects are displayed in Fig. 16.2 as a function of list length.

If the effect of list length on mean latency were the consequence of a few outliers, then the high quantiles should increase steeply and the low quantiles very little. Instead, all the statistics increase at respectable rates; slopes of fitted linear functions are provided in the figure caption. Even the minimum, often an unstable statistic, increases reliably with list length, at a rate of 8.4 ± 2.8 msec/word ($p < .05$). Slopes of the fitted functions do change systematically with p-value, indicating that the effect of list length on latency cannot be described merely as translation of its distribution with no change in shape. Instead, the increasing separation among quantiles with increasing list length reflects the existence of a modest increase in latency variability as well. (The mean standard deviation of the latency distribution increases from about 22 msec for $n = 1$ to 28 msec for $n = 5$, in the present experiment.)

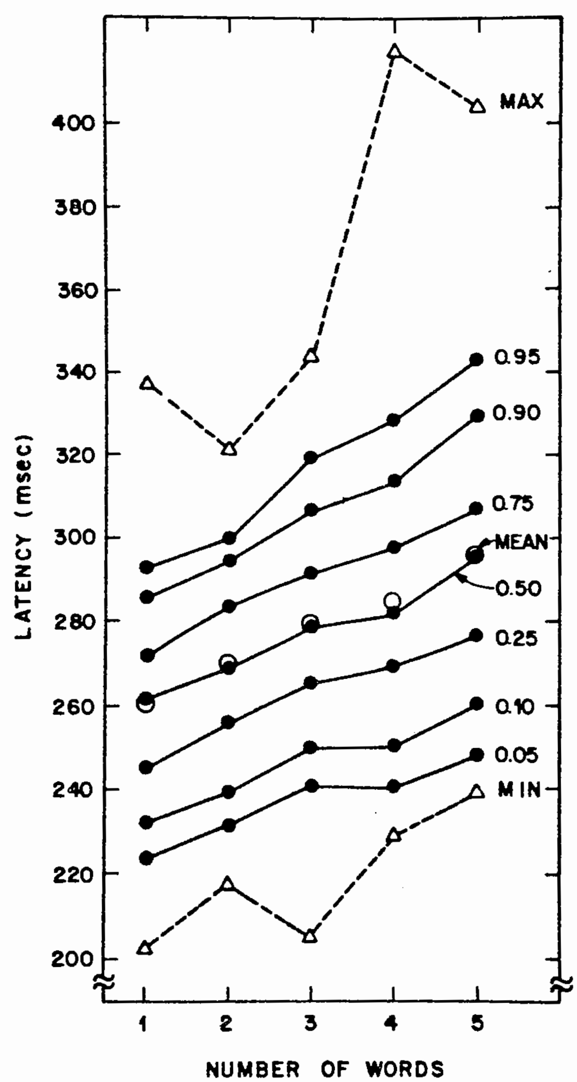

FIG. 16.2. Mean latency quantiles, mean latency extremes, and mean latency for lists of one to five words in the weekdays experiment. Each statistic was determined or estimated from a distribution of about 60 observations from each subject; values were then averaged over the four subjects. Quantiles (with p-value indicated) are shown by filled circles, minima and maxima are shown by triangles, and mean latencies are indicated by open circles. Slopes of linear functions fitted to each of these statistics, in msec/word, are: minimum, 8.4; mean, 8.7; maximum, 22.9; quantiles, from lowest to highest, 5.7, 6.7, 7.6, 8.1, 8.5, 10.6, 12.8.

V. TEST OF A PHYSIOLOGICAL HYPOTHESIS ABOUT THE LATENCY EFFECT BY MEASUREMENT OF INITIAL FUNDAMENTAL FREQUENCY

Our favored explanation of the effects of the length of an utterance on its latency and duration is the subprogram-retrieval version of the sequence-preparation hypothesis (LD, III).[6] This account, as well as the alternatives we presented in LD and rejected in its favor, attributes the effects to mental processes (e.g., subprogram retrieval) rather than to physiological or mechanical constraints on the articulators. One basis for our not having considered this latter class of peripheral physiological explanations was our assumption of *element invariance* (LD, I).

As it applies to the initial element (word) in an utterance, the strongest form of the element-invariance assumption asserts that no characteristics of this first element (other than when it is produced) are systematically influenced by sequence length. A weaker but sufficient version of the assumption (and one that is less likely to fail) asserts that sequence length has no systematic influence on the delay between the hypothetical motor commands for the initial element and the acoustic events that determine its measured latency. (If this assumption were to fail, then sequence length could influence utterance latency indirectly, through its direct effects on this delay.) In our early experiments, however, we included no direct tests of the element-invariance assumption; arguments for and against the assumption were adduced only from the degree of regularity of the latency function (LD, I; LD, VI).

We were encouraged to test the assumption more directly by Kenneth N. Stevens. He proposed an interesting physiological explanation of the latency effect which depends on the assumption being violated by the initial word in an utterance.[7]

According to Stevens' proposal, (1) before an utterance is initiated the laryngeal muscles must be tensed; (2) the longer the utterance, the greater the initial tension; and (3) more time is required for greater tension to be achieved from a resting state. (Utterance "length" here would presumably be measured in stress-group units, as discussed in LD, IV). Furthermore, (4) greater initial tension should produce a higher fundamental voice frequency (F_0) at the start of the utterance. It is this last part of

[6]According to this hypothesis, a program consisting of a set of n subprograms, one for each response unit in the utterance, is constructed prior to the signal. Then, before each unit is executed, its subprogram is retrieved. The retrieval process is one (such as sequential search) whose mean duration increases linearly with n. The effect of n on retrieval duration for the first subprogram produces the latency effect, and the sum of the effects of n on retrieval durations for the $n - 1$ remaining subprograms produces the duration effect.

[7]Stevens' proposal was made as a personal communication at the present Symposium.

Stevens' proposal that permits it to be tested in a straightforward way: In contrast to our element-invariance assumption, the proposal implies that the F_0 contour of the initial word of a longer utterance should be higher, or at least start higher.[8]

Sorenson and Cooper (Chap. 13, Table 13.5) give us some reason for expecting an increase in initial F_0 with utterance length. Under their perhaps more "natural" conditions for speech production, peak F_0 in the first stressed syllable in long sentences (13.8 words) was about 13.5 Hz (or 6%) higher than in short sentences (7.5 words). But there are at least three reasons why this finding may not apply to our situation: We used random words, not sentences; our range of list lengths is below their range of sentence lengths; and our subjects were encouraged to complete their utterances as rapidly as possible, unlike theirs. Furthermore, the first stressed syllable in Sorenson and Cooper's sentences was preceded by an unstressed syllable: Suppose that pitch contours (or, alternatively, F_0 targets for potential stressed syllables) were at equal heights at the initiation of sentences of different lengths, and declined more rapidly for shorter sentences (which they have been found to do; see references in Section VI.) Then the initial F_0 measure reported by Sorenson and Cooper would vary with utterance length, even though the actual F_0 associated with the initial unstressed syllable (or, alternatively, the F_0 target associated with an initial stressed syllable if one had existed) remained constant.

Given Stevens' proposal together with our observed effect of about 10 msec per stress group on utterance latency, how large an effect of utterance length on initial F_0 would we expect? One rough indication of the time difference needed to achieve different degrees of tension of the laryngeal muscles *prior* to voicing is perhaps provided by data on the relation between the size of an increase in F_0 *during* actual voicing and the minimum time to achieve that increase. (Such evidence can provide only an indication, however, since maximum rates of tension change may not be equal under voiced and silent conditions.) Sundberg (1979) and Ohala and Ewan (1973; data reported in Sundberg, 1979) have shown that the additional time needed for an F_0 increase of 12 semitones relative to 4 semitones is no more than about 25 msec, or about 3 msec per semitone.[9] (One semitone corresponds to an increase of 6% in frequency.) For a

[8]See Sorenson and Cooper (Chap. 13) for a discussion of F_0 and its measurement.

[9]One possible objection to regarding this kind of measurement as relevant to our problem is that insofar as F_0 changes are similar to aimed arm displacements, the time to make a change will depend on the required endpoint precision (which appears not to have been examined in detail in Sundberg's study, 1979) as well as the size of the required change (Fitts & Peterson, 1964).

latency increment of 40 msec (approximately the difference between utterances of $n = 1$ and $n = 5$ words) to be attributable to the time required to achieve a higher initial F_0, the difference between initial F_0 values would therefore have to be at least 13 semitones, or about an octave.

Lists in the experiment we used to test Stevens' proposal contained from $n = 1$ to $n = 5$ elements drawn from a vocabulary of five monosyllabic words: *bee, cow, day, pie,* and *toe*. We chose these words (all starting with stop consonants) so that word boundaries would be clearly indicated by minima in the amplitude envelope; the primary aim of the experiment was to permit us to measure the interword time intervals as a function of serial position (discussed in Sec. VII). We used two types of list, *homogeneous* lists, containing repetitions of the same word, and *heterogeneous* lists, containing n distinct words. For both types and for each length, each of the five words appeared equally often in each serial position. We used a new group of four female speakers. Each list appeared on three successive trials; the utterances subjected to the F_0 analysis include one list of each type with each starting word, drawn from one of the last two of these trials, (i.e., 10 utterances for each list length and each of the four subjects).

The procedure on each trial of the present experiment, as well as of the experiments to be described in Sections VIII and IX, was different from the procedure described in LD, II: After an auditory warning sounded and a visual fixation pattern appeared, the list of words was presented in a column. It remained visible until the end of the response. Except on "catch" trials, the reaction signal (a tone-burst) occurred $4.4 + 0.6n$ sec after display onset (where $1 \leq n \leq 5$ is the list length). The reaction signal was preceded at 0.7-sec intervals by two noise bursts that functioned as "countdown" signals. Subjects were instructed to complete each utterance as soon as possible after the reaction signal, consistent with high accuracy.

In conjunction with an LPC (linear predictive coding) analysis of each utterance, we computed an amplitude value for each 10 msec segment, as well as an F_0 value for each segment in which voicing had occurred.[10] Word boundaries were defined by minima in the resulting amplitude functions.

The experiment produced qualitatively typical latency effects for both homogeneous and heterogeneous lists; the slopes of a line fitted by least squares to the mean data are 14.0 ± 3.1 msec/word and 16.1 ± 2.8 msec/

[10]See Rabiner, Cheng, Rosenberg, & McGonegal (1976), Section II F, for a description of the pitch-detection algorithm, due to B. S. Atal; see Atal & Hanauer (1971) for a comprehensive discussion of LPC analysis of speech.

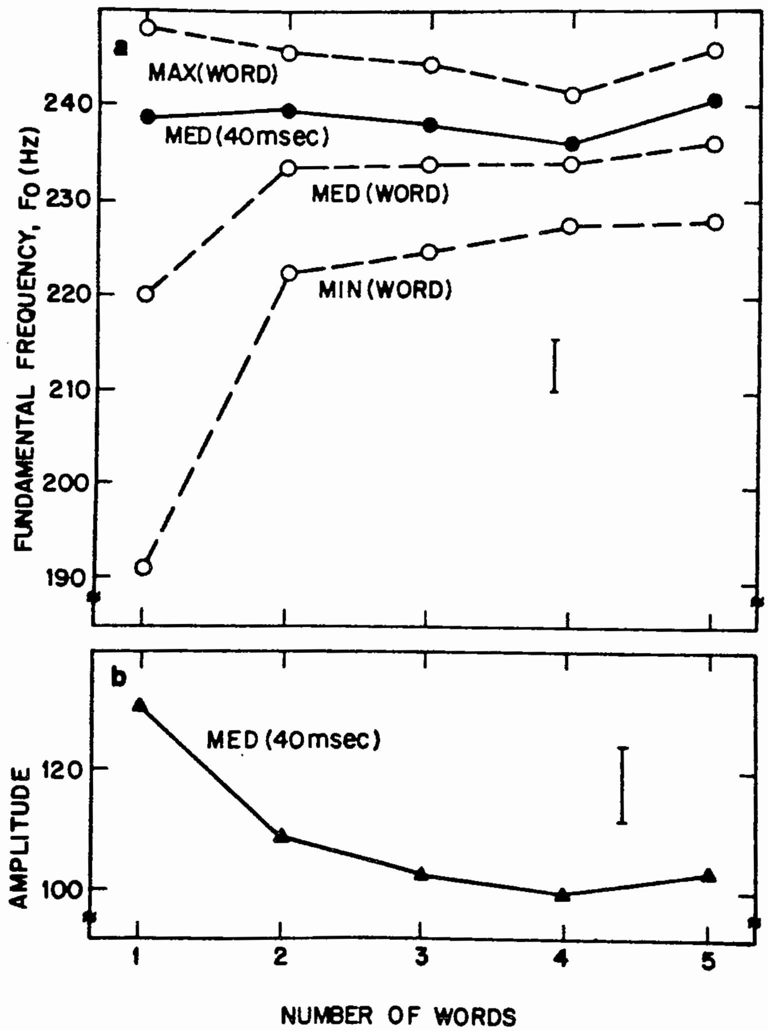

FIG. 16.3. Acoustic measures of the initial word in utterances of one to five monosyllabic words. Values are means of the indicated statistics over homogeneous and heterogeneous lists, five different initial words, and four subjects; 40 observations per point. Also shown are estimates of $\pm SE$ of these means, based on *subjects* × *list length* interactions. a. Fundamental frequency (F_0) in Hz is shown by circles: Open circles show maxima, medians, and minima over all 10-msec segments in voiced portion of first word; filled circles show medians over first four such segments (40 msec). b. Median amplitude over first four voiced segments, in arbitrary units.

word, respectively, somewhat larger values than we had obtained in earlier experiments with different subjects.

Acoustic measures of the first word are shown as a function of utterance length in Fig. 16.3. The measure that is most appropriate for testing Stevens' proposal is the fundamental frequency at the start of the first word; we have used the median of the four F_0 values obtained from the first four segments in each utterance where F_0 was defined (i.e., where voicing had occurred). As can be seen from the figure (filled circles), rather than increasing with list length this statistic is essentially invariant; the slope of a line fitted to the five points is 0.02 ± 1.85 Hz/word. The maximum F_0 over all voiced segments in the word (whose mean is within 1 Hz of the maximum over the first four such segments) can also be seen to be essentially invariant; the slope of a fitted line is −0.87 ± 1.65 Hz/word. These findings support the element-invariance assumption and require us to reject Stevens' proposal.

The other statistics do show effects of list length, however. Minimum F_0 and median F_0 during the voiced portion of the initial word are both significantly lower ($p < .01$) when $n = 1$ than when $n \geqslant 2$; differences among lengths $n \geqslant 2$ are not statistically significant. This observation is probably best thought of in relation to the fact that for $n = 1$ the first word is also the last, and therefore participates in a terminal fall in F_0. (See Sec. VI.) Similarly, the initial amplitude of the utterance is significantly higher ($p<.05$) when $n = 1$ than when $n \geqslant 2$; again, differences among lengths $n \geqslant 2$ are not statistically significant. The existence of distinctive frequency and amplitude characteristics for isolated words reinforces the reservation we expressed in LD, I as to whether single elements would satisfy the invariance assumption.

VI. EFFECTS OF UTTERANCE LENGTH AND SERIAL POSITION ON FUNDAMENTAL FREQUENCY: THE DECLINATION EFFECT IN RAPID SPEECH

The tendency for F_0 to drift downward from the beginning to the end of a natural speech utterance (or "intonation group" within an utterance) appears to characterize most languages (Bolinger, 1978; Ohala, 1978; Pierrehumbert, 1979). Measurements of isolated sentences in American English show the amount of declination to be 20–40 Hz and to be relatively independent of utterance length (e.g., Maeda, 1976; Sorenson & Cooper, Chap. 13). The approximate invariance of declination with length implies that the *rate* of declination varies inversely with length. Thus, a characteristic of the whole utterance (its length) is reflected in a local feature (rate of F_0 declination). This observation has been taken as evidence for the advance planning of the entire utterance, just as has our observation that, under conditions of time pressure, sequence length influences interelement time. (See Sec. VII.)

The similarity of these arguments encouraged our interest in whether the declination effect would appear under the conditions of our experiments. Here utterances were of unstructured lists (rather than phrases or sentences) which subjects were instructed and trained to complete as rapidly as possible; subjects were also asked to avoid phrasing, to stress each word equally, and to speak "in a monotone." The presence of F_0 declination might add to the evidence for advance planning, and would encourage us to feel that our subjects' utterances were not as "unnatural" as our experimental conditions might lead one to expect. Systematic differences in F_0 as a function of serial position and length would also bear on the validity of the element-invariance assumption.

The experiment and measurement procedure as well as the set of data used in the analysis were described in Section V. For the present analysis we determined, for each word, the median F_0 value over all its 10-msec voiced segments. Results, averaged over subjects and over all lists of each length, are shown in Fig. 16.4, together with the MED(40 msec) measure of initial F_0 from Fig. 16.3.

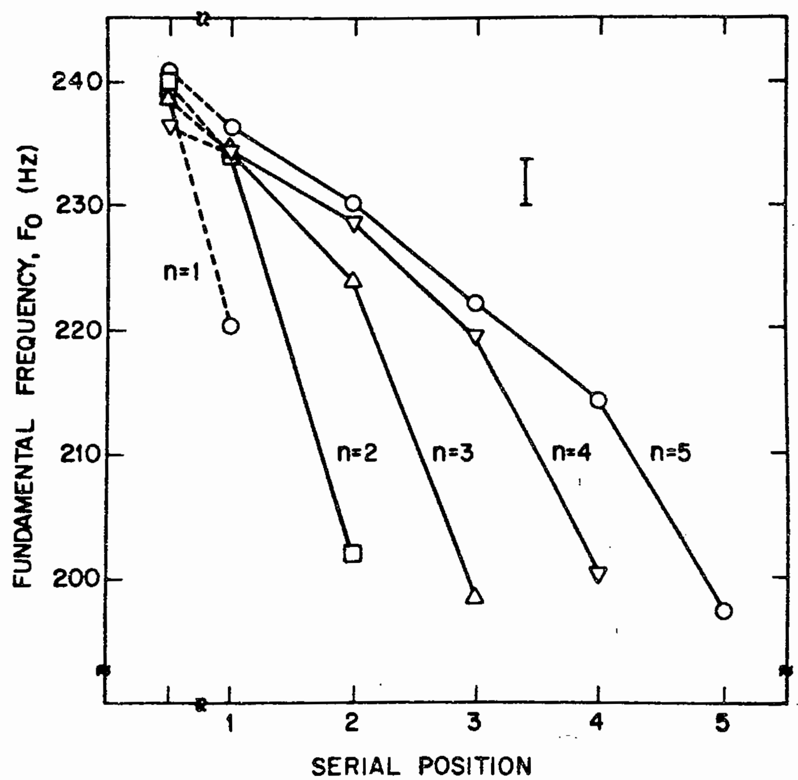

FIG. 16.4 Fundamental frequency (F_0) in Hz as a function of serial position in lists of lengths $n = 1$ to $n = 5$. For the leftmost points (no serial position indicated), median values of F_0 were taken over the first four 10-msec segments in the voiced portion of each initial word; points represent the means of these medians for each length. For other points, median values of F_0 were taken over *all* voiced segments in each word; points represent the means of these medians for each length and serial position. Also shown is an estimate of $\pm SE$ of the displayed means, based on the *subjects* $\times$ *serial-position* interaction. Means were taken over two types of list, five words, and four subjects; 40 observations per point. The mean time interval corresponding to a separation of one serial position was 143 msec.

Averaged over lengths $n \geqslant 2$, the mean declination, measured from initial F_0 (leftmost points in the figure), is large and significant: 39.5 ± 6.7 Hz. As has been reported for "natural" speech, the total drop in F_0 appears not to depend systematically on length; for $n = 2, 3, 4$, and 5, the mean declination values are 38.0, 40.1, 36.3, and 43.5 Hz, respectively. Excluding the final word at each length (whose median F_0 is affected by the terminal fall) and averaging over lengths $n \geqslant 3$, the effect is smaller but still highly significant: 19.4 ± 2.7 Hz. If the effect is measured over all serial positions (for $n \geqslant 2$) the *rate* of declination (F_0 change per serial position) declines significantly as n increases; if the final word is excluded, although the rate decreases with n, this decrease cannot be shown to be statistically significant with tests based on between-subject variability. However, taken together, these results suggest that the main features of the pitch declination phenomenon can be found in rapidly spoken lists. This finding adds to the evidence for advance planning, and indicates that even under the special conditions of our experiments, utterances display an important organizational property of natural speech.

VII. THE EFFECT OF SERIAL POSITION ON THE INTERWORD INTERVAL

One aspect of the findings reported in LD that we consider relatively important is the similarity of the duration functions for speech and typewriting.[11] Both are well approximated by quadratic functions, which indicates that the mean interval between successive elements increases linearly with number of elements. (See LD, Fig. 15.6a, b.)

In typewriting, because individual interstroke intervals are easily measured, we were also able to determine the mean interstroke interval for each serial position (LD, Fig. 15.6c). We found a dominance relation among serial-position functions: For each serial position (counting backward from the last interstroke interval in the series, or forward from the first), the duration of the mean interstroke interval increased with n. Furthermore, the serial-position functions were found to be concave downward, and indicated that for the longer sequences the stroke rate is slower toward the middle of the sequence than at either end. We felt it was important to examine further the generality of these findings by determining analogous serial-position functions for rapid speech.

[11]Sequence length also produces increases in typing latency. The duration and latency effects are therefore both found in motor performances that differ radically in training history, muscular complexity, measurement method, and the amount of time pressure and relative speed outside the laboratory. Given such generality, potential measurement artifacts are of less concern, and the underlying process is more likely to be fundamental.

The experiment described in Section V was designed to permit using the amplitude envelope of each utterance to determine the mean intervals between successive response units: We used words beginning with stop consonants and we balanced word identities across serial positions. However, identification of the "beginning" of the execution of a response unit embedded in continuous speech presents difficulties of definition as well as measurement, and such an effort depends on relatively arbitrary decisions. Furthermore, since in terms of our model (LD, VII) our aim is to measure events that indicate the time intervals between the beginnings of the *command stages* for successive units, these difficulties are compounded by our assumption that the execution of a unit can be prolonged beyond the termination of its command stage (to continue in parallel with the *retrieval* and *unpacking* stages for the next unit). In attempting to deal with these difficulties we made two assumptions. First, we assumed that each monosyllabic word is a response unit; second, we assumed that the occurrence time of an event relatively early in the word (the transition between the stop consonant and the vowel, which was associated with an abrupt rise in the amplitude envelope) would adequately measure (on the average, and to within an additive constant) the starting time of the command stage for the unit in that serial position.

We digitized subjects' speech at 10kHz and determined the mean absolute sample value (an amplitude measure) in adjacent 10-msec intervals. Linear interpolation between successive means then defined an amplitude envelope that usually showed a single peak for each word in the utterance and a single trough between each word and the next. On 86% of the correct trials we were able to choose a criterion amplitude value that intersected the ascending flanks of exactly n amplitude peaks. We chose an algorithm to do this that produced mean amplitude criteria that depended relatively little on n. The time differences between successive intersections then defined a series of $n - 1$ intervals between the "beginnings" of successive words. (We also established that these interword intervals were relatively independent of the value of the amplitude criterion—an important check, given the crudeness of the segmentation procedure.)

Results are shown separately in Fig. 16.5 for homogeneous and heterogeneous lists, in the same format as the typing data (LD, Fig. 15.6c) to facilitate comparison. The same dominance relation found in typing also appears in the speech data: At each serial position the mean interword interval increases with n, and with the exception of one value (homogeneous lists of length $n = 2$), the duration of the first interword interval increases with n. As in typing, therefore, a characteristic of the entire sequence—its length—influences the initiation delay of each of its constituents. The existence of such an effect supports the idea that execu-

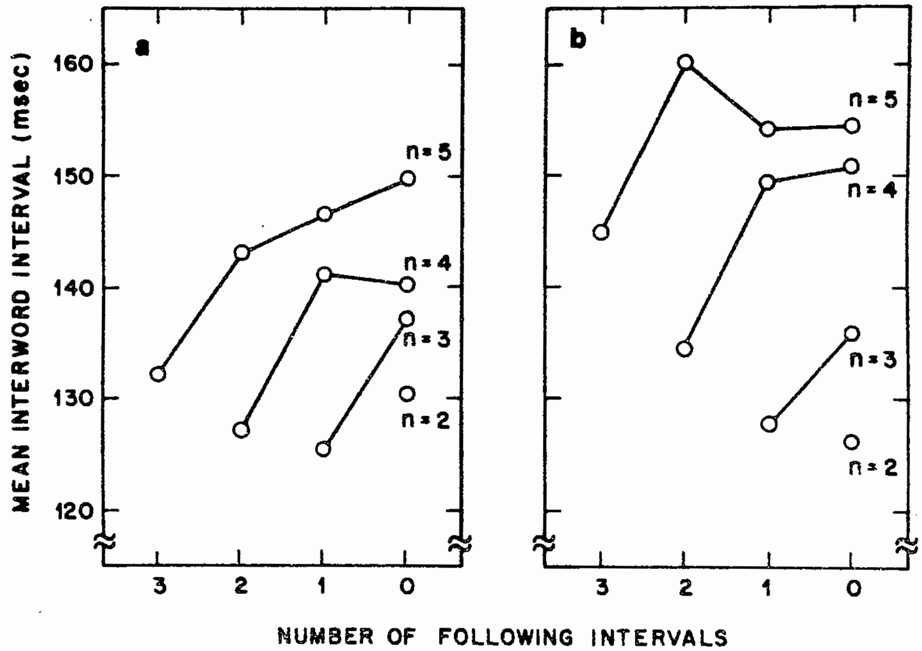

FIG. 16.5. Mean interword interval for each serial position in lists of monosyllabic words of lengths $n = 2, 3, 4$, and 5. Results are averaged over four subjects; about 50 observations per point. a. Homogeneous lists. b. Heterogeneous lists. Intervals early in a list have more intervals following them and appear toward the left of the plots.

tion of the constituents is based on a representation of the entire sequence, and that at a level relevant to the control of timing, the entire sequence is represented during execution of each of its elements.

However, the average rate at which the interelement interval increases with serial position in speech is greater than in typing: Rapid utterances tend to slow down, and the final acceleration found in typed sequences is absent or much attenuated in the speech data. In terms of our subprogram-retrieval model (LD, VII), these differences in the effect of serial position on the interval between elements in speech and typing, in combination with the same (linear) effect of length on the *mean* interval, could be attributed to different average search *orders* in the context of the same mechanism of self-terminating search.

The separation between serial-position curves is smaller for homogeneous than heterogeneous lists, reflecting the fact that the slope of the mean rate function (or the quadratic coefficient of the duration function) is significantly smaller: Estimated coefficients are 4.7 and 14.3 msec/word^2 for homogeneous and heterogeneous lists, respectively.[12]

[12]Taken together with the substantial difference between *rate-function* slopes for homogeneous and heterogeneous lists, the similarity of the shapes of their *serial-position functions* challenges the assumption embodied in our model (LD, VII) that the same retrieval mechanism underlies both functions. Further evidence against the assumption appears in Section VIII. A second dissociation of effects that challenges our present model is suggested by the finding in this experiment that rate-function slopes can differ between two kinds of lists, without a corresponding difference in latency-function slopes (values in Sec. V).

This feature of the data from repeated monsyllabic words contrasts with our findings in the weekdays experiment (LD, II); there we found that neither the latency nor duration functions for lists composed of repetitions of the same weekday name were especially unusual. We have not yet identified the conditions under which homogeneous and heterogeneous lists are associated with parametrically distinct latency or duration functions.

VIII. LOCALIZATION WITHIN WORDS OF THE EFFECTS OF UTTERANCE LENGTH AND SERIAL POSITION

We have seen that under the conditions of our experiments the time per word increases with the length of an utterance and also tends to increase with serial position within an utterance. Since the utterance can be regarded as a stream of articulatory gestures partitioned into subsequences, each associated with a response unit, it is tempting to try to determine the loci within these subsequences of the length and serial-position effects: To what extent are these effects concentrated in specific parts of each response unit?

One possible answer is that the effects are not localized, but instead result from relatively global differences in the rates at which articulatory gestures occur. In its simplest form, however, our model (LD, VII) provides a very different answer. According to the model, utterances depend on a series of *command* stages, one for each response unit, which are interleaved with *retrieval* and *unpacking* stages. Utterance length influences the duration of only the retrieval stage, and does this by influencing the mean duration of a search process. Furthermore, the serial-position effects are produced by variation across serial positions in the duration of this same search process. Effects of both length and serial position are therefore restricted to the time intervals between the end of one command stage and the beginning of the next, and timing *within* command stages should not be affected. If the *execution* of a response unit coincided with its command stage, the consequence of this argument would be straightforward: There would be silent intervals between one unit and the next, and any effects of length or serial position would be localized exclusively within these intervals.

We have had to elaborate our model, however (LD, VII): To reconcile the continuity of speech with the temporal discreteness of the hypothesized sequence of command stages we assumed that the execution of a unit could be prolonged beyond the termination of its command stage, so as to continue in parallel with the search and unpacking stages associated with the next unit. To force utterance duration to be suitably sensitive to the durations of retrieval stages, we also had to assume that the portion of the response unit that was prolonged in this way could be adjusted or truncated so as not to delay the next unit. The elaborated model suggests that although the duration of an entire response unit may depend on utterance length and serial position, this may not be true for the timing of within-unit features that occur relatively early in the unit, because they depend on direct rather than delayed control from the command sequence and may not be subject to adjustment or truncation.[13]

We felt that a plausible initial test of the elaborated model would be provided by measuring utterances whose response units were two-syllable words. As a starting assumption, we conjectured that for each such response unit the command sequence would be in direct control until at least the beginning of the second syllable. Our model then implies that the time intervals between the beginnings of successive syllables in an utterance consisting of a sequence of such disyllables should not be equally sensitive to the effects of utterance length and serial position. Instead, these effects should be restricted to those intervals that include speech events from the beginning of the second syllable of one word to the beginning of the first syllable of the next. We shall call these "across-word" intervals, and the others "within-word" intervals.

To perform the test we constructed lists from a vocabulary of disyllables with primary stress on the first syllable. (Examples of the words we used are *copper, dagger, pebble,* and *token*.) Each word contained two stop consonants, one at the beginning of each syllable; all of the other constituent sounds were vowels or sonorants. The articulatory pattern of each word was therefore *closed* (C_1), *open* (V_1), *closed* (C_2), *open* (V_2). A transition between stop consonant and vowel or sonorant involves the release of the stop and the onset of voicing and is marked by an abrupt rise in the amplitude envelope. For each word, therefore, the amplitude envelope usually contained two distinct peaks, and the time between any two successive rises in amplitude provided a measure of an intersyllabic interval. The resulting decomposition of a three-word utterance is shown schematically in Fig. 16.6.[14] Word identities were balanced over lengths

[13]An alternative elaboration of the model, with different consequences, would postulate advance knowledge, as early as the beginning of the execution of unit k, of the duration of the unpacking and retrieval stages for unit $k + 1$. This information would be used to adjust the rate at which the articulatory gestures were executed throughout unit k, so as to allow them to fill the available time from one command stage to the next. This alternative seems less appealing, however, not only because of its greater complexity, but because it reduces the plausibility of a retrieval process being required to determine unit $k + 1$ after execution of unit k has begun.

[14]For expository clarity we have assumed that in the words used in our experiment, the second consonant, C_2 is included within the second syllable rather than the first. Syllabication of such words is a matter of controversy, however. Fortunately for our purposes, the transition between the stop and the following vowel or sonorant is near the beginning of the syllable, under any of the accepted definitions.

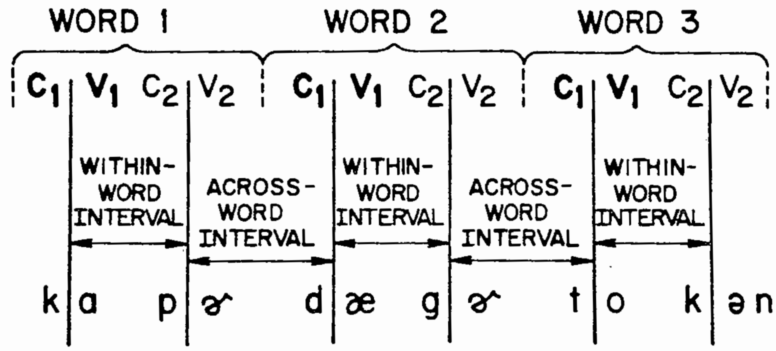

FIG. 16.6. Schematic representation of the decomposition of a three-word utterance. Each word can be regarded as a stressed syllable (designated (C_1V_1) followed by an unstressed syllable (designated C_2V_2), where each syllable consists of a single stop-consonant (C_i) followed by a vowel and/or sonorant (V_i). Each vertical line represents an abrupt rise in the amplitude envelope at the transition from C_i to V_i, which intersects an amplitude criterion that is fixed for the entire utterance. Time intervals between successive intersection points correspond to within-word and across-word intervals, as shown. The decomposition of *copper, dagger, token* is shown as an example.

and serial positions. Lists were either heterogeneous ($1 \leqslant n \leqslant 4$) or homogeneous ($1 \leqslant n \leqslant 5$). Four subjects were run for two days, providing a total of about 200 trials per list length for each type of list. Our experimental procedure is described in Section V; our measurement methods were as in Section VII, with one exception: given an utterance of n words, we searched for a criterion amplitude value that intersected the ascending flanks of exactly $2n$ amplitude peaks, so as to define a series of time intervals between successive syllables.

It has to be recognized that this method of segmenting an utterance is relatively crude and may provide measures that distort the time intervals between the significant underlying articulatory events. We are particularly concerned about the potential distorting effects of amplitude variations, either between or within utterances. However, interpreted with proper caution, the method seemed suitable for an initial attempt to localize the effects of length and serial position, and clearcut results would help to justify it.

Results, averaged over the four subjects, are shown in Tables 16.1 and 16.2, with estimates of $\pm SE$ based on between-subject variability. In Table 16.1, under "within words" are shown measures of the effect of list length on the mean time interval from the amplitude rise of the first syllable in each word to the amplitude rise of the second. Durations of these intervals were averaged over serial positions; the resulting means were then regarded as a function of length and fitted by a straight line. Tabulated values are the slopes of such lines, averaged over subjects. Under "across words" is shown a corresponding measure derived from the time interval from the amplitude rise of the second syllable in each

TABLE 16.1
Rate of Increase with List Length, n, of Mean Time Interval between Successive Syllables (msec/word)

	Within Words ($n \geq 1$)	*Across Words* ($n \geq 2$)	*Difference*
Homogeneous lists ($n \leq 5$)	0.9±1.6	4.6±0.9**	3.7±2.3
Heterogeneous lists ($n \leq 4$)	2.7±1.8	12.6±2.7**	9.9±3.8*

*$p < .05$, **$p < .01$.

word (excluding the last) to the amplitude rise of the first syllable of the next word. (The sum of the two measures estimates the rate at which the total time per word increases with utterance length—the slope of the rate function. The effect of type of list on this sum—15.3 msec/word versus 5.5 msec/word for heterogeneous and homogeneous lists, respectively—is similar to the effect found for lists of monosyllables, which we discussed in Sec. VII.)

Table 16.1 shows that for both kinds of list less than 20% of the overall effect of utterance length on production rate is reflected in our measure of within-word events. This is so despite the fact that the mean within-word interval (132.1 msec) was longer than the mean across-word interval (106.7 msec). That the length effect is almost entirely restricted to the later parts of each response unit supports the elaborated subprogram-retrieval model (and argues against any alternative model that would produce a global effect of length on articulation rate).

The effects of serial position on the within-word and across-word intervals are shown in Table 16.2. For each length, the duration of the relevant interval was regarded as a function of serial position and fitted by a straight line. (A linear function was not an unreasonable representation.) The slopes of these lines were then averaged over lengths to obtain the tabulated values. (The sum of the within-word and across-word measures estimates the rate at which the total time per word increases with serial position. The difference between these sums for the two kinds of list—11.8 msec/position versus 10.9 msec/position for heterogeneous and homogeneous lists, respectively—is very small, indicating similarity among serial-position functions, just as we found for lists of monosyllables.[15])

Table 16.2 shows that for both kinds of list almost all of the effect of serial position on the total time per word is reflected in our measure of within-word events—a result that clearly violates expectations based on our model. Thus, although the within-word interval is essentially inde-

[15]See footnote 12 on p. 522.

TABLE 16.2
Rate of Increase with Serial Position of Mean Time Interval between Successive Syllables, Averaged over List Lengths, n (msec/position)

	Within Words	*Across Words*	*Difference*
Homogeneous lists ($2 \leq n \leq 5$)	$8.7 \pm 1.8^{**}$	2.2 ± 1.1	-6.5 ± 2.8
Heterogeneous lists ($2 \leq n \leq 4$)	$13.2 \pm 1.8^{**}$	-1.4 ± 2.7	$-14.6 \pm 3.3^{**}$

$^{**}p < .01$

pendent of utterance length when *averaged* over serial positions (Table 16.1), it does depend strongly on serial position (Table 16.2). In contrast, the across-word interval is virtually independent of serial position (Table 16.2) but increases with length (Table 16.1). Thus the effects of length and serial position appear to have different loci in the stream of articulatory gestures. This double dissociation puts into further question the idea, embodied in our model, that the two effects are manifestations of the same retrieval process. If these findings can be confirmed with more refined measurements, we shall have to modify our model so as to provide an alternative account of the effect of serial position.

In one possible modification of the model, the serial-position effect is caused by planned variation of rates specified by *command* stages (represented in the corresponding subprograms) rather than with variation from unit to unit in *retrieval*-stage durations. Suppose that at the command level such variation influenced the representations of articulatory gestures *throughout* the response unit, rather than being arbitrarily restricted. Now consider the hypothesized influence (described earlier) of the start of the next command stage on termination time for the ongoing execution process. This influence (which mediates the length effect) would mask or dominate any effects of the proposed command-level variation on *later* parts of each execution process. Serial-position effects would then be restricted to early parts of each execution process, as observed, and the double dissociation would follow.[16]

[16] Any mechanism that generates the effect of serial position on the interword interval must be made consistent with the form of the length effect—i.e. with the observed linearity with length of the *mean* interword interval. Such consistency can follow naturally if both length and position effects are generated by the same self-terminating search process, as initially proposed. But for the modified serial-position mechanism, consistency requires a separate property: The mean (over positions) of the amount of lengthening it produces must be approximately independent of (or linear with) utterance length.

Consideration of these alternative serial-position mechanisms points up an important issue that bears on any programming model: the locus of effects that extend across units (also exemplified by coarticulation and pitch declination). Such effects may be embodied directly in the series of subprograms or, alternatively, may be regarded as emerging at "run" time as consequences, e.g., of interactions among successive execution processes.

IX. THE TIMING OF UTTERANCES COMPOSED OF WORDS VERSUS NONWORDS, AND THE ROLE OF LEXICAL MEMORY IN RAPID SPEECH

There has been much study of the overall process of transforming print into speech in naming tasks, and of the effects of prior experience with the words or letter strings named (e.g., Smith & Spoehr, 1974; Theios & Muise, 1977). But attempts to investigate separately the response-production mechanism in the naming of individual words have thus far yielded little information about how speech production itself interacts with and makes use of the stored information about words in long-term lexical memory. In this section we consider the role of lexical memory in the production of sequences of one or more words.

According to our model for the production of rapid prespecified utterances (LD, VII), the subject generates a motor program that contains one subprogram for each response *unit* in the utterance, where the unit is the *stress group* (LD, IV). Furthermore, the retrieval process responsible for the effects of utterance length on latency and duration operates at the level of these subprograms. The hypothesized process consists of a search through a buffer containing the set of subprograms that were assembled and loaded in advance of the reaction signal. For the experiments reported in the present chapter, each *word* is a stress-group unit; for utterances composed of lists of monosyllabic *nonwords* produced with the same stress pattern as word lists, each nonword is a unit.

Our aim in comparing words with nonwords was to contrast two kinds of units. For words, preexisting articulatory routines (which have perhaps evolved to be particularly efficient) may be stored in lexical memory. Nonwords, however, cannot have preexisting representations. Suppose such stored articulatory routines exist for words.[17] One possible role of lexical memory in speech production would then be to serve as a library of detailed word-production information that would be retrieved at execution time. In terms of our model, when the subprogram for a unit was executed, it would fetch the word routines required for its stress group from lexical memory. Information in the buffer would therefore control the sequence of lexical memory retrievals, but the buffer would not itself

---

[17]In another linguistic domain, Terzuolo and Viviani (1979) have argued from regularities in the timing of keystrokes that professional typists make use of stored routines ("motor engrams") for the typing of common words. On the other hand, studies of word and nonword naming suggest that phonological representations of words may often be "computed," rather than directly retrieved from memory. (See, e.g., Glushko, in press.) While this would not preclude the use of whole-word articulatory routines (accessed by the word's phonological representation rather than its "identity"), it would make such use less plausible.

contain all the information needed to produce the utterance. Production of an utterance containing a list of nonwords, for which lexical memory contained no routines, would require a qualitatively different mechanism; for word and nonword lists we would expect qualitatively different patterns of latencies and durations.

A second possible role of lexical memory would again be to furnish preexisting word routines, but here they would be retrieved and incorporated prior to program execution in the subprograms stored in the buffer. Although it might take more time to *assemble* the program for a list of nonwords, the same process of program *execution* would be used for nonword lists as for word lists, and we would expect a qualitatively similar pattern of latencies and durations. However, because the newly produced routines for nonwords might be less efficient than the preexisting routines for words, execution of the program might be slower, with consequent effects on parameters of the duration and/or latency functions. A third possibility is that lexical memory plays a minimal role in speech production. Even though an articulatory unit (the stress group) appears to control utterance timing, preexisting articulatory routines at the word level would be used in neither assembling nor executing the motor program. Instead, phonological representations of the words (or nonwords) in the utterance would first be obtained, by retrieval from lexical memory or by application of grapheme-phoneme rules, for example. These phonological representations would then be converted into articulatory routines for stress-group units either when the program was assembled or when it was executed. The conversion (which might use preexisting stored articulatory routines for lower-level phonic *subunits*) would be carried out afresh for each new utterance, whether it was composed of words or nonwords: we would expect to find no production-timing differences between the two kinds of utterance.

To compare the production of words and nonwords, while avoiding effects that might be due to accidental or systematic differences among their constituents, it seemed critical to use words and nonwords that were phonologically matched. The existence of coarticulation effects implied that not only should the sets of constituent phonemes be the same for words and nonwords, but also at least the set of pairs of adjacent phonemes, both within items and across item boundaries. We achieved this matching requirement by searching a phonetic dictionary for sets of phoneme triples containing four words and four nonwords related as shown in Table 16.3.

We used four such matched sets in the experiment, arranging them in either word lists or nonword lists that contained one to four different items. Subjects were instructed to pronounce the two kinds of list with the same stress pattern, attempting to stress each item equally. Different

TABLE 16.3
Phonologically-Matched Words and Nonwords

Examples		*Symbolic Description*[a]	
Words	*Nonwords*	*Words*	*Nonwords*
vote	vate	(1, 1, 1)	(1, 2, 1)
hone	hane	(2, 1, 2)	(2, 2, 2)
vain	vone	(1, 2, 2)	(1, 1, 2)
hate	hote	(2, 2, 1)	(2, 1, 1)

[a] The three ordered digits in the symbolic representation correspond respectively to the first, second, and third phoneme, each being drawn from a (different) pair of alternatives. For example, if F_1 and F_2 are alternative first phonemes, S_1 and S_2, second, and T_1 and T_2, third, then (2, 1, 2), which is *hone* in our examole, corresponds to $F_2S_1T_2$.

lists were used on successive trials. For the data to be reported, we used the experimental procedure described in Section V.[18]

Latency and duration data are shown in Fig. 16.7. The latency functions are almost identical and are approximately linear, with the fitted lines accounting for 98.0% and 97.1% of the variance for words and nonwords, respectively; deviations from linearity were not statistically significant. Averaged over lengths, the latency difference is well within the experimental error: Mean latency for nonwords is 0.2 ± 2.5 msec greater than for words.

The mean duration data are fitted remarkably well by quadratic functions (even considering that only one degree of freedom remains after fitting): The fitted functions account for more than 99.9999% of the variance for both words and nonwords. However, the fitted quadratic functions differ by about 2 msec in both constant and linear coefficients; averaging over lengths, the duration per item for nonwords is 2.4 ± 1.0 msec greater than for words, a very small but statistically significant difference. (Given the mean duration per word of 142.5 msec, this difference represents only 1.7%.) Averaging over word and nonword lists, the mean slope of the latency function (15.0 ± 4.7 msec/word) and the mean quadratic coefficient of the duration function (17.3 ± 2.0 msec/word^2) differ little, as usual, but are somewhat larger than in other exper-

[18]For the present analysis we classified one of our eight subjects as an outlier, based on her performance with word lists. Even with these lists she was both unusually slow and unusually variable: Her duration mean and standard deviation were greater than the means of these statistics over the remaining seven subjects by 4.7 and 7.4 *SD*s, respectively.

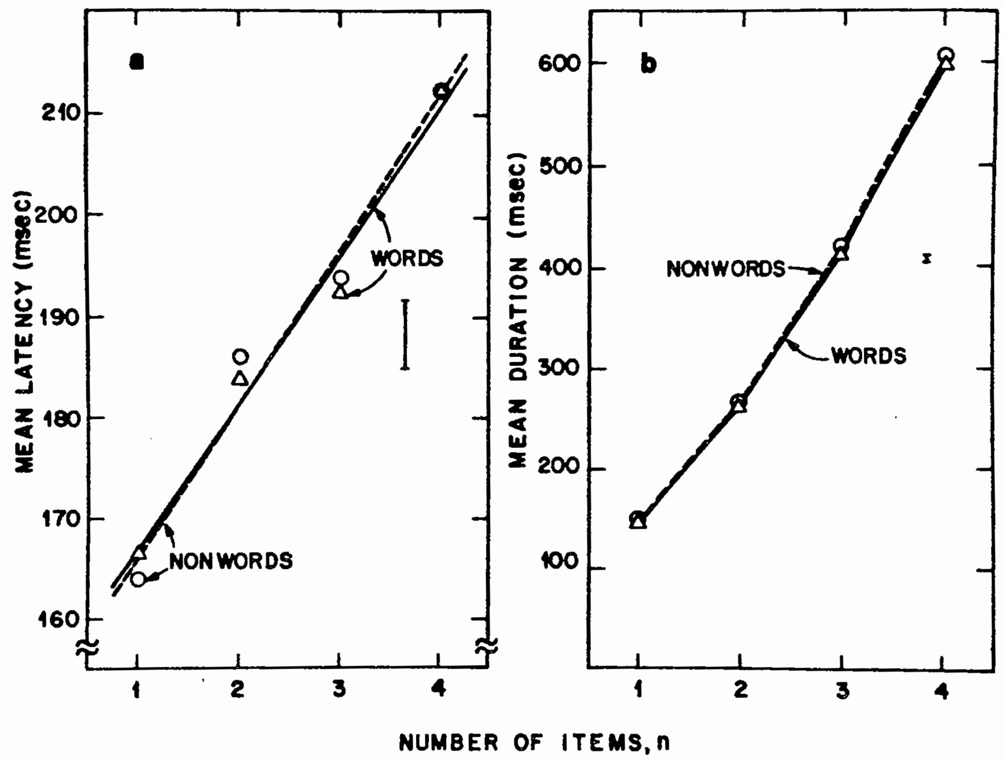

FIG. 16.7. Experiment comparing words and nonwords: Results are averaged over seven subjects; about 140 observations per point. Triangles and circles show results with words and nonwords, respectively. a. Mean latencies, estimate of $\pm SE$, and fitted linear functions [words, $166.8 + 14.6(n - 1)$ msec; nonwords, $166.1 + 15.3(n; - 1)$ msec]. b. Mean durations, estimate of $\pm SE$, and fitted quadratic functions [words, $144.3 + 99.5(n - 1) + 17.1(n - 1)^2$; nonwords $146.3 + 101.4(n - 1) + 17.4(n - 1)^2$]. See LD, Table 15.1 for fitting procedure. *SE* estimates in this figure were chosen to be appropriate for evaluating word–nonword differences. They are based on the *item-type* × *length* × *subjects* interactions in analyses of variance.

iments, probably because of sampling differences between groups of subjects.[19]

Spoken under time pressure, then, prespecified utterances composed of nonwords have latency and rate characteristics that are qualitatively the same and quantitatively almost the same as utterances composed of words. Since our method of phonological matching extended only to pairs

---

[19]In this experiment each set of words and nonwords was used to construct lists presented to subjects over a series of trials. Thus, on the average test trial on which the data displayed in Fig. 16.7 are based, subjects had had about 23 previous opportunities to pronounce each item. (Indeed, the trials that provided approximately the first eight opportunities per item were excluded from the displayed data.) Although unlikely, it seemed possible that even these few minutes of experience might invest nonwords with the properties of words.

In one evaluation of this possibility we analyzed sets of *initial* trials on which subjects of adjacent phonemes whereas coarticulation effects are known to have a greater range (see, e.g., Kent & Minifie, 1977), it is even possible that the small duration difference we did find is a consequence of imperfect phonological matching.

Our results indicate that during the latent interval and the execution of prespecified utterances no reference is made to information stored in lexical memory; if any previously stored word-level articulatory routines are used at all, they are apparently incorporated into the motor program before the utterance begins.[20] Furthermore, if preexisting articulatory routines for word units exist and are used in assembling the program, their efficiency advantage over newly produced routines, in terms of execution speed, appears to be negligible. These results also support two other claims we have made: First (LD, III), the latency effect is not caused by variations in the load on ordinary short-term memory. And second (LD, IV), the response unit underlying performance in our experiments is articulatory (such as the stress group) rather than semantic (such as the word).[21]

---

had had an average of seven (and as few as zero) opportunities to pronounce each item. For these trials, the mean latency difference between words and nonwords is as small as in the main body of data, and not statistically significant; the mean duration difference per item, although not significantly different from zero, is twice as large for these trials as in the main data (4.8 ± 5.1 msec rather than 2.4 ± 1.0 msec).

In a second test of word-nonword equivalence we compared error rates for lists of words versus nonwords in the main body of data. The overall error rate was low (2.50% of trials, composed of 1.70% "pronunciation errors" and 0.80% "memory errors"). If stored articulatory routines were used during production of words, and such routines could be constructed for nonwords after a few trials, then we would expect pronunciation errors to be equally frequent for words and nonwords. Instead, most (68%) of the trials with pronunciation errors were nonword trials. When the lists were displayed only briefly, after comparable familiarization, the percentage of trials with pronunciation errors increased from 1.70% to 4.30%. Because this increase was restricted to nonwords, even more (83%) of these errors were now associated with the nonwords.

These subsidiary findings suggest that the remarkable similarity of the latency and duration functions for words and nonwords is not due to the existence of stored articulatory routines for the nonwords that developed during the experiment.

[20]Further support for this conclusion is provided by comparisons of articulation rates and latencies for lists of nouns drawn either from distinct semantic categories (*n*-category lists) or from one category. It is well known that both recognition and naming of printed words are facilitated by the existence of a semantic relation between successive words; the effect is as large as 50–100 msec per word. (See, e.g., Meyer & Schvaneveldt, 1976.) Furthermore, such effects have been explained in terms of the spread of excitation through a semantic network stored in long-term memory, which lowers word-detector thresholds. If production of prespecified utterances involved access to information stored in lexical memory, one might expect semantic relatedness to produce facilitation of production as well. (One argument against the facilitation of pronunciation per se when the utterance is *not* specified well in advance is that the facilitation of word naming—which does require overt pronunciation—is

ACKNOWLEDGMENTS

We are indebted to O. Fujimura, M. Y. Liberman, N. H. Macdonald, D. L. Scarborough, and K. N. Stevens for helpful discussions and advice, to A. S. Coriell, W. J. Kropfl, and M. J. Melchner for technical support, and to J. C. Johnston, D. L. Noreen, and J. S. Perkell for comments on the manuscript.

REFERENCES

Atal, B. S., & Hanauer, S. L. Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America*, 1971, *50*, 637–655.

Bolinger, D. Intonation across languages. In J. H. Greenberg, (Ed.), *Universals of human language: Volume 2, Phonology*. Stanford, Calif.: Stanford University Press, 1978.

Conrad, B., & Schönle, P. Speech and respiration. *Archiv für Psychiatrie und Nervenkrankheiten*, 1979, *226*, 251–268.

Fitts, P. M., & Peterson, J. R. Information capacity of discrete motor responses. *Journal of Experimental Psychology*, 1964, *67*, 103–112.

Glushko, R. J. The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, In Press.

Kent, R. D., & Minifie, F. D. Coarticulation in recent speech production models. *Journal of Phonetics*, 1977, *5*, 115–133.

Maeda, S. *A characterization of American English intonation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, Mass., 1976.

Meyer, D. E., & Schvaneveldt, R. W. Meaning, memory structure, and mental processes. *Science*, 1976, *192*, 27–33.

---

not significantly greater than the facilitation of word recognition—which does not; see Meyer & Schvaneveldt, 1976.) One-category lists should then be articulated substantially more rapidly than n-category lists. This appears not to be the case: In a pilot experiment with five subjects we constructed lists of the two kinds, of length $1 \leq n \leq 5$, from sets of five monosyllabic nouns drawn from the categories of animals, fruits, drinks, body parts, and house parts. Differences in latencies and rates were negligible and not significant. Averaged over lengths, mean latency for n-category lists was only 1.2 ± 5.3 msec greater than for 1-category lists; mean duration per word was 3.7 ± 3.7 msec greater, a difference of only 1.8%.

[21]The word-nonword experiment can be regarded as a demonstration that sequence familiarity at the level of phonemes or articulatory gestures has no effect on utterance timing. Evidence favoring an effect of sequence familiarity at the level of words (or stress groups), which complicates matters, is found in the experiment using lists of weekday names (LD, II). For lists of length $2 \leq n \leq 5$ the mean duration per word was 190.9 msec and 208.1 msec, for "normal" and random sequences, respectively; the difference of 17.2 ± 4.2 msec, or 9.0%, is highly significant. Furthermore, the effect seems to be associated with retrieval rather than command stages in our model, since it is localized in the quadratic rather than the linear coefficient of the duration function. (See caption of LD, Fig. 15.2.) One explanation (not yet tested) that is consistent with such specificity and that would avoid the need to modify our model is that familiar word sequences occasionally combine to form higher-order response units such that a stress group incorporates more than one word; this should be revealed through stress-pattern alterations.

Ohala, J. J. Production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey.* New York: Academic Press, 1978.

Ohala, J., & Ewan, W. G., Speed of pitch change. *Journal of the Acoustical Society of America,* 1973, *53,* 345.

Pierrehumbert, J. The perception of fundamental frequency declination. *Journal of the Acoustical Society of America,* 1979, *66,* 363–369.

Rabiner, L. R., Cheng, M. J., Rosenberg, A. E., & McGonegal, C. A. A comparative performance study of several pitch detection algorithms. *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 1976, *ASSP-24,* 399–418.

Smith, E. E., & Spoehr, K. T. The perception of printed English: A theoretical perspective. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II. Acta Psychologica,* 1969, *30,* 276–315.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning.* New York: Academic Press, 1978.

Sundberg, J. Maximum speed of pitch change in singers and untrained subjects. *Journal of Phonetics,* 1979, *7,* 71–79.

Terzuolo, C. A., & Viviani, P. About the central representation of learned motor patterns. In R. Talbot & D. R. Humphrey (Eds.), *Posture and movement: Perspectives for integrating neurophysiological research in sensory-motor systems.* New York: Raven Press, 1979.

Theios, J., & Muise, J. G. The word identification process in reading. In H. J. Castellan, Jr., D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (Vol. 2). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.