6

The Latency and Duration of Rapid Movement Sequences: Comparisons of Speech and Typewriting

Saul Sternberg
Stephen Monsell
Ronald L. Knoll
Charles E. Wright

Information Processing in
Motor Control and Learning

ISBN 0-12-665960-5

I. Introduction

We communicate verbal information in three principal ways: We write, we speak, and we use keyboards. Maximum information rates in speech and typing are higher than in writing (Seibel, 1972), and as more of our work uses computers, keyboard entry is becoming increasingly important. This chapter reports some new findings about the temporal patterns of rapid movement sequences in speech and typewriting and what these patterns might mean in relation to the advance planning or "motor programming" of such sequences. We shall be concerned with how response factors affect the time to initiate a prespecified rapid movement sequence after a signal (the "simple-reaction" time) when the goal is to complete the sequence as quickly as possible, as well as how such factors affect the rate at which movements in the sequence are produced.[1] The response factor of central interest will be the number of elements in the sequence.

Most existing research on skilled performance has been concerned with *perceptual*–motor skills. For example, in most studies of typing behavior, performance has been measured when new segments of text are read concurrently with the typing of segments that have just been read (see, e.g., Butsch, 1932; Shaffer, 1976). Nonetheless, it has been argued that a major source of errors in typing is in the control of finger movements rather than in the perception of what is to be typed (see, e.g., Shaffer & Hardwick, 1969; Van Nes, 1976). Partly for this reason and partly because we believe it is desirable to study aspects of skilled performance in isolation, we have tried in these experiments on speech and typewriting to study movement processes uncontaminated by the concurrent perception of new material.

These studies began with two accidental findings (reported in Monsell & Sternberg, Note 1). The first was that the number of words in a brief rapid

---

[1] Some similar questions about handwriting are addressed in Chapter 7.

utterance influenced the time to initiate the utterance, even though the talker knew what he would have to say well in advance of the reaction signal. This finding seemed surprising, particularly in view of the claim based on previous studies (Eriksen, Pollack, & Montague, 1970; Klapp, 1971, 1976) that the latency (or reaction time) for saying a single word, known in advance, is not affected by the number of syllables it contains. The second finding was that the functions relating the duration of these rapid utterances to the number of words they contained were concave upward rather than being linear, indicating that words in longer sequences were produced at slower rates.

Our interest in the effect of the length of a movement sequence on its latency was based partly on the possibility that it reflects a latency component used for advance planning of the entire sequence: The length effect would then measure the extra time required to prepare extra elements. The idea that changes in reaction time might reflect changes in *sequence preparation* in this way seems to have been first proposed by Henry and Rogers (1960), who found that simple-reaction time increased with the number of elements in a sequence of movements made with one arm. According to their model, part of the reaction time includes the time to gain access to stored information concerning the whole sequence: a process akin to loading a program into a motor buffer, with sequences containing more elements requiring larger programs, and larger programs requiring more loading time. Numerous studies have since been made of effects of factors such as the extent, duration, and "complexity" of arm and hand movements on their latency. Both simple-reaction and choice-reaction paradigms have been used. However, as indicated in the reviews of this work by Hayes and Marteniuk (1976) and by Kerr (Chapter 3 of this volume), both paradigms have given rise to conflicting and controversial findings. We consider briefly three issues that are relevant to these conflicts and to the experimental methods we decided to adopt.

A. Element Invariance and the Measurement of Response Complexity

One issue that arises in examining previous work is the proper index of response complexity. An increase in the extent or duration of a movement sequence may not necessarily increase the amount of planning required, as measured by the number of "instructions" or "subprograms" in a "program." (Once this issue is raised, it suggests an explanation for the contrast mentioned above: The latency of an utterance could be influenced by the number of words in the utterance but not by the number of syllables in a word if the "programming unit" were a suprasyllabic sequence such as a word or a stress group, rather than something smaller such as a syllable or an articulatory gesture.)

The aspect of response complexity we have manipulated in our experiments is the number of elements (spoken words, keystrokes) in a movement sequence, which we varied over a wide range (from one to five) relative to previous studies. To minimize the probability that effects of the number of elements in a sequence on its latency or duration are trivial consequences of differences in the elements themselves, we have tried to ensure that the elements are as fixed as possible, regardless of the sequence in which they are embedded. Insofar as the approximation to this *element-invariance requirement* is a good one, the wide range enables us to study the form of the function relating latency to the number of elements and to investigate the relation between the "programming units" and the elements we manipulate. In addition, insofar as the form of the function is simple (e.g., linear), the data would suggest that the invariance requirement is approximately satisfied.

Before leaving this issue it is helpful to be somewhat more precise about the requirement of element invariance. Let us divide elements in a sequence into four classes: interior elements (*i*) both preceded and followed by other elements, beginning elements (*b*), terminating elements (*t*), and single elements (*s*) neither preceded nor followed by others. Now consider the elements that are present in sequences of increasing length: *s, bt, bit, biit, biiit*. To study the form of the function relating latency (or duration) to the number of equivalent elements in a sequence we can restrict ourselves to the *changes* in performance that result from equivalent *increases* in length, starting from a short sequence.

Suppose that we start with sequences of length $n = 2$. Then all increases in length can be regarded as resulting from the addition of *i*-elements. We therefore need to assume that *i*-elements are all equivalent. In addition, since elements already present must not change as we increase sequence length, all *b*-elements must be assumed to be equivalent regardless of sequence length, and similarly for *t*-elements. There is no need for *b*- and *t*-elements to be equivalent to *i*-elements or to each other.

If we start with sequences of length $n = 1$ rather than $n = 2$, however, we must also assume the equivalence of *s*- and *b*-elements and of *i*- and *t*-elements (or, alternatively, of *s*- and *t*-elements and of *i*- and *b*-elements). One can imagine this requirement not being met, especially given the evidence, from speech, of bidirectional coarticulation effects (see Kent & Minifie, 1977). Insofar as our data cause us to question this equivalence, we must restrict our attention to the performance functions for $n \geqslant 2$.

B. Response Effects in Simple Reactions versus Choice Reactions

A second issue raised by previous work is whether choice reaction or simple reaction is the more appropriate paradigm for investigating the effects of the characteristics of a response on the planning of that response. One argument (e.g., Klapp, 1976) assumes that after sufficient practice one movement sequence (simple reaction) can be fully prepared in advance of the signal, but that more than one (choice reaction) cannot; therefore, choice-reaction latencies are more likely to reflect response-planning operations. However, there is little independent evidence favoring either of the assumptions. Certain response variations have been observed to produce larger effects on latency of the choice reaction, and this fact has been taken as evidence that advance planning between signal and response plays a larger role there than in the simple reaction. However, this inference depends on the idea that the only locus of response-factor effects is response planning; the inference is not justified if response factors also influence additional operations that might be required for choice reactions but not for simple reactions, such as "translation" from stimulus to response or "response selection."

There are two kinds of "compatibility" effects in the choice-reaction paradigm that suggest the influence of response factors on operations other than response planning. Consider first the finding that the latency of a particular response to a particular stimulus depends on the mappings of the other possible responses on the other possible stimuli (e.g., Duncan, 1977). This suggests that changes in latency produced by varying the attributes of a specified response to a specified stimulus are likely to depend on S–R mappings of other pairs—a dependence that might be hard to explain solely in terms of the planning of that response.

Consider second the effects of compatibility of entire stimulus and response ensembles ("SE–RE compatibility," as contrasted with the "S–R compatibility" of mappings of the same stimulus and response ensembles; Brainard, Irby, Fitts, & Alluisi, 1962): The change in latency induced by switching from one R-ensemble to another depends on the S-ensemble. This indicates that at least one of the processing stages between stimulus and response is influenced by the identities of both (Sternberg, 1969a, Section 5.3); hence, any effect of response factors might arise at least in part in that stage and could therefore depend on (interact with) stimulus factors—another dependence that might be hard to explain in terms of response planning.

Considerations like these not only weaken the argument just mentioned for using the choice-reaction paradigm, but also show that the effects of response (or stimulus) factors cannot be assigned conclusively to response (or stimulus) processes without suitable control experiments. While this caveat applies to simple- as well as choice-reaction paradigms, we believe the hazards to be greater in the choice situation, where it is likely that a more complex series of processes determines the latency and that stimulus–response interactions are larger. Indeed, in those studies we know of where the same variations in S–R mapping were examined in both choice-reaction and simple-reaction paradigms, S–R compatibility effects that were substantial in the former proved to be vanishingly small in the latter (Callan, Klisz, & Parsons, 1974; Anzola, Bertoloni, Buchtel, & Rizzolatti, 1977).

Another consideration that arises in the choice-reaction paradigm is the possibility of effects of *response–response compatibility*. Evidence has been accumulating that the other possible responses in an experiment influence the latency of a specified response, not merely by virtue of their number, but also their kind. Thus, in the case of binary choice, the latency of a response is shorter if it is paired with another that could be performed at the same time (Berlyne, 1957), or that is performed by a finger on the opposite hand (Kornblum, 1965), or that involves a movement in the same direction (Megaw, 1972), or that contains the same initial phoneme (Sanders, 1970). Given the available evidence, these effects cannot be unequivocally distinguished from S–R or SE–RE compatibility and conclusively assigned to response processes. But insofar as similar effects are observed with arbitrary stimulus ensembles and arbitrary S–R mappings, an interpretation in terms of response processes is compelling. We believe that ultimately an account of the planning and execution of responses will have to explain R–R compatibility.[2] For the present, however, such considerations complicate the interpretation of response effects: In the choice-reaction paradigm, the influence of altering one of the responses on the latency of that response could depend on the identity of the other (not performed) response.

Our approach, then, has been to study the effects of sequence length on latency in the simple-reaction paradigm; we believe that contributions to these effects from stimulus discrimination and S–R translation processes are minimized in that paradigm, and it permits us to defer the issue of R–R compatibility. If under these conditions there remains an orderly dependence of latency on the nature and number of elements in the entire movement sequence, this dependence would seem particularly worth investigation, since there would appear to be fewer alternatives that compete with the hypothesis of advance planning of the entire sequence than in the choice-reaction paradigm. However, we do not wish to argue that just because we use the simple-reaction paradigm, any effects of response factors on latency can be immediately assigned to a response-planning process that occurs after the signal but before the start of the response. It is still necessary to pit this hypothesis against the promising alternatives that remain, such as delays associated with operations that are required to maintain a description of the response in short-term verbal memory, failure of the invariance requirement discussed above, or effects of sequence length on the time to retrieve the first element from a previously loaded motor-program buffer. Indeed, as we shall attempt to show, the results we have obtained thus far tend to favor a somewhat different hypothesis from the one that first attracted us to the problem.

---

[2] A study by Rosenbaum (Note 2) can be regarded as a first attempt to provide a process model of R–R compatibility effects. He has argued, with experimental support, that under some conditions only the movement "features" not shared by the pair of responses in a binary choice-reaction task are prepared after the signal, whereas the features they do share are prepared before the signal. Note that "R–R compatibility" has usually been applied to simultaneously performed responses.

C. Relation between Advance Planning and Feedback Control

A third issue raised by previous work, and one that dominates much writing on motor processes, is the relation between the advance planning of a movement sequence and the influence of feedback during the execution of that sequence. Clearly the existence of any sensory delay requires that brief movement elements be controlled independently of the peripheral feedback they produce (e.g., Welford, 1974). The idea of a central motor program for the control of entire sequences seems to have arisen from the observation that sensory delays were too great to permit feedback ("closed loop") control even from one element to the next, in rapid performance (e.g., Lashley, 1951; but see also Adams, 1976).

The "program" concept has been restricted in recent years by the idea that the *only* way organisms deal with limited feedback delays in executing rapid movements is to preplan entire sequences (e.g., Schmidt, 1972), rather than, for example, planning some movements concurrently with the execution of earlier movement. Thus, in his influential definition, Keele (1968; see also Russell, 1976) proposed that "a motor program may be viewed as a set of muscle commands that are structured before a movement sequence begins, and that allows the entire sequence to be carried out uninfluenced by peripheral feedback [p. 387]."

This definition has seemed to suggest to some investigators that advance planning generates only command sequences that can be executed without any feedback ("open loop"), rather than, for example, programs that include instructions for sensing and responding to feedback, programs that can themselves be altered in response to feedback, or even programs that consist of ordered sets of "response images" (e.g., Greenwald, 1970; Adams, 1976) to which feedback from the movement sequence is compared.

We believe that it is inappropriate to restrict the "program" concept to cases of sequence control without feedback. Suppose that for a particular kind of movement sequence we had a hierarchical analysis in terms of sequences of units, each consisting of a sequence of subunits, and so forth. At each level of the hierarchy, control would have to be exercised over the selection, sequencing, and timing of the subunits, as well as over other attributes. At each level of the hierarchy and for each attribute, separate and largely independent questions could be raised, first about the roles of central and sensory sources of feedback and second, about the time relations between preparation and execution. Possible roles of feedback include, for example, serving as a cue that triggers the onset of the next subunit in a sequence, or providing information used in an error-correction process. Possibilities for the scheduling of preparation range from preparing each subunit after the previous one has been executed, through preparing later subunits while earlier ones are being executed, to preparing the whole sequence in advance.

We feel that questions about the existence and extent of advance planning are separable from questions about the precise role played by feedback, and we suspect that the methods appropriate for answering them are very different. In our view the experiments to be described in this chapter bear principally on the issue of the time relation between planning and execution.[3]

II. Experiments on Speech

With minor deviations the procedure on each trial of the three speech experiments to be described was as follows: First a short list of digits or words was presented sequentially and visually, at a rate of about 1 sec per item. The lengths and compositions of the lists were varied from trial to trial by means of a different balanced randomization for each subject. The list was followed by a fixed delay of about 4 sec that subjects could use for rehearsing silently and preparing to respond. On about 85% of the trials a visual "recite signal" (an illuminated rectangle) appeared at the end of the delay. This signal was preceded at 1-sec intervals by two brief "countdown" signals (the first signal auditory, and the second visual), which were included to minimize the subject's time uncertainty about when he might be required to respond. On the remaining 15% of the trials the recite signal was omitted; subjects were not to respond on these "catch trials," which were included so as to prevent anticipations on signal trials.

Instructions, feedback, scores, and cash bonuses were designed to encourage subjects to *complete* the reciting of each list as soon as possible after the signal, while maintaining a low error rate. (The error rates were in fact negligible and will not be discussed.) The subjects were four female high school students who were well practiced in experiments requiring rapid reciting.

Using an energy-sensitive speech detector with a low threshold, we made two measures of each response: its *latency*, measured from signal onset to the start of the utterance, and its *duration*. (The subjects were attempting to minimize the sum of these two measures.) Each subject had about 200 trials per day.

A. Numbers in Ascending Sequence

In our initial studies we had used lists of randomly ordered letters or digits. We later observed the same effects with well-learned sequences that place a minimal load on memory. Thus, in the first experiment presented here the lists were subsequences of one to five items drawn from the natural number sequence

---

[3] The experiments on speech were selected from a larger series described in Monsell and Sternberg (Note 1) and to be reported in greater detail elsewhere by Monsell and Sternberg. The experiment on typewriting was selected from a series to be reported in greater detail elsewhere by Sternberg, Knoll, and Wright.

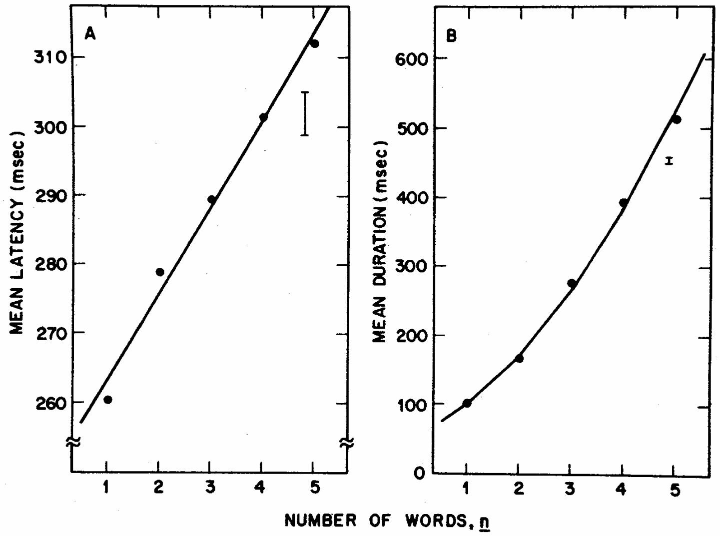

Figure 1 Ascending numbers experiment: The results are averaged over four subjects and over starting numbers; about 140 observations per point. (A) Mean latencies, estimate of standard error (±*SE*), and fitted linear function (Table I). (B) Mean durations, estimate of ±*SE*, and fitted quadratic function (Table I). Note the difference between the ordinate scales. See Table I, Footnote *a*, for the fitting procedures. The *SE* estimates in Figure 1A–6B were chosen to be appropriate for revealing badness of fit. They are therefore based on mean squares for high-order interactions from analyses of variance in which *subjects* was treated as a fixed effect. Table I appears on p. 139.

1, 2, . . . , 9, and starting with one of the five numbers 1, 2, . . . , 5. On the single day of testing there were about seven trials per subject for each of the 25 possible lists.[4]

Figure 1A shows that mean latency increased approximately linearly with the number of words (n) in the list at a rate of 12.6 msec/word. Thus, the time to start saying *two–three–four–five–six*, for example, was about 50 msec greater than the time to start saying *two*. Latency functions were similar across starting digits (slopes of the fitted functions are 10.2, 10.1, 15.3, 14.0, and 13.3 msec/word for lists starting with 1, 2, 3, 4, and 5, repectively) and across subjects (slopes are 10.1, 9.5, 19.4, and 11.3 msec/word for the four subjects). Linear regression

---

[4] A feature that distinguishes this experiment from our other experiments is that, although beginning items are balanced in lists of different lengths, populations of interior and terminating items differ systematically across lengths, thereby violating the element-invariance requirement of Section I,A. This could bias the results somewhat, especially for duration data.

accounts for 98.7% of the variance among mean latencies; deviations from linearity were not statistically significant.[5]

Figure 1B shows that the increase of duration with list length is distinctly nonlinear; the quadratic function shown fits well however, accounting for 99.8% of the variance among mean durations.[6] Acceleration of the duration function implies that the average articulation rate depends on list length: the longer a list, the greater the average time from the beginning of one word to the beginning of the next (this idea is made precise in Section V,A). Subjects responded well to the request that they complete their utterances rapidly: Their average articulation rate of about 9.4 words/sec is high relative to previously reported maximum rates (Hudgins & Stetson, 1937; Landauer, 1962).

B. Weekdays in Normal, Random, and Repeating Sequence

In a second experiment we compared subjects' production of three kinds of weekday sequences: normal (e.g., *Wednesday–Thursday–Friday–Saturday*), random, without replacement (e.g., *Monday–Friday–Wednesday–Sunday*), and repeating (e.g., *Monday–Monday–Monday–Monday*). The cyclic structure of the days of the week allowed us simultaneously to match populations of beginning words, terminating words, and interior words across list lengths (even in the familiar normally ordered lists) and across conditions. Lists contained from one to five words, and the two days of the experiment provided about 25 trials per length per condition per subject.

Mean latency (Figure 2A) again increased approximately linearly with list length. Slopes of the fitted linear functions do not differ significantly across conditions. The average latency slope is actually *smaller* for weekdays (8.8 msec/word) than for digits (12.6 msec/word) despite their greater syllabic length, but this between-experiment difference is not reliable.

Despite the similarity of latency functions, durations were, of course, much greater for lists of weekdays than for lists containing the same number of digits.

---

[5] To test whether the latency effect could result from a small but increasing proportion of extreme observations with longer lists, we examined the latency distributions. Mean quantiles for each list length and mean standard deviations were obtained by averaging separate estimates from latency distributions for each subject and starting digit. The slopes of linear functions fitted to the mean 20 and 80% points of the distributions were 10.2 and 14.1 msec/word, respectively, indicating that the entire distribution, and not simply the upper tail, is affected by list length. The greater rate of change for the higher quantile reflects a small increase in dispersion with list length; mean standard deviations of latencies increase from about 22 msec ($n = 1$) to 30 msec ($n = 5$).

[6] We chose a quadratic function partly because of its success in describing other duration data for $1 < n < 5$. Although departures from the fitted quadratic function are small, they are statistically significant in this experiment (but not in the others to be reported). See Footnote 4 for a possible explanation of this difference between experiments.

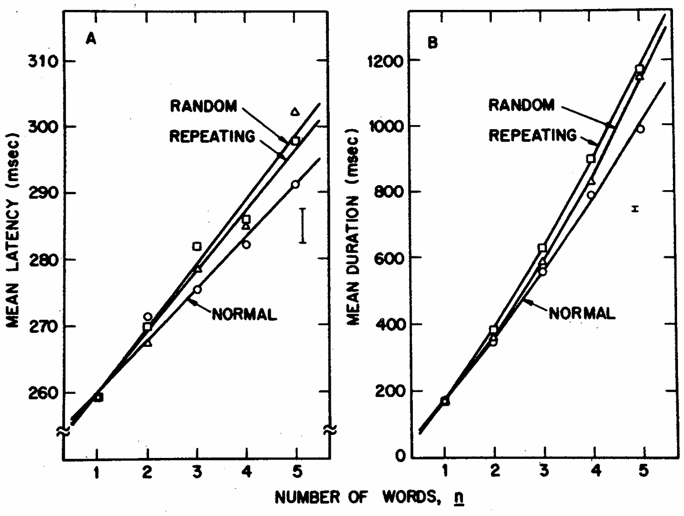

Figure 2 Weekdays in normal, random, and repeating sequences: The results are averaged over four subjects; about 100 observations per point. (A) Mean latencies, estimate of $\pm SE$, and fitted linear functions: normal, $259.9 + 7.8(n-1)$; random, $259.9 + 9.6(n-1)$; and repeating, $259.9 + 9.1(n-1)$. (B) Mean durations, estimate of $\pm SE$, and fitted quadratic functions: normal, $168.8 + 174.1(n-1) + 8.6(n-1)^2$; random, $168.8 + 174.0(n-1) + 16.6(n-1)^2$; and repeating, $168.8 + 201.2(n-1) + 12.7(n-1)^2$. The three fitted functions in each panel were constrained to pass through a common fitted value at $n = 1$.

Mean durations (Figure 2B) are significantly nonlinear, but very well described by quadratic functions[7]; these functions differ significantly across conditions.

For understanding our results, the relation between the repeating condition and the others will be of particular interest. One way to make this comparison is to subtract from each coefficient of the fitted function for the repeating condition the mean of the corresponding coefficients for normal and random conditions. The resulting differences are 0.4 ± 0.9 msec/word for the slope of the latency function, 0.1 ± 4.0 msec/word^2 for the quadratic coefficient of the duration function, and a significant 27.2 ± 3.2 msec/word for the linear coefficient of the duration function.[8] The only reliable effect of constructing an utterance from repetitions of the same word rather than from distinct words is to increase the linear coefficient of the duration function.

---

[7] The percentages of variance among the mean durations accounted for by fitted quadratic functions are 99.92, 99.96, and 99.98% for normal, random, and repeating conditions, respectively, and 99.99% for the means over conditions.

[8] When quantities are stated in the form $a \pm b$, b is an estimate of the standard error of a, based on between-subject variability.

When we compare the duration function for numbers to the average function for the weekdays conditions (see Table I) we find that the large duration difference is localized primarily in the constants (101.1 versus 168.8 msec) and linear coefficients (57.6 versus 183.1 msec/word). The quadratic coefficients (12.2 versus 12.6 msec/word^2) are almost identical.

C. Words of One and Two Syllables

Because they were produced in separate experiments, comparison of the data from number and weekday lists could only suggest whether and how the coefficients of the latency and duration functions depend on number of syllables per word. For more precise estimation of these effects we ran an experiment that was larger and that incorporated deliberate variation of word length.

Lists varied in length from one to four words; the words in a list were either all one-syllable words or all two-syllable words. The vocabulary consisted of 72 common nouns and was constructed so that all the two-syllable words were stressed on the first syllable and contained (an approximation to) one of the one-syllable words as the first syllable. (Examples of such embedded pairs are *bay–baby, rum–rumble, track–tractor, cow–coward,* and *limb–limit.*) Our aim was to bring our manipulation as close as possible to the addition of an unstressed syllable to a given stressed syllable. On each of the eight days of the experiment

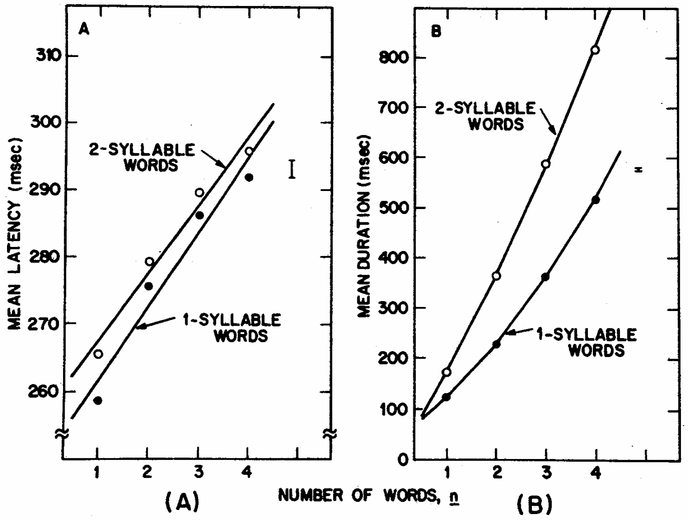

Figure 3 Experiment comparing words of one and two syllables: The results are averaged over four subjects and 8 days; about 400 observations per point. (A) Mean latencies, estimate of $\pm SE$, and fitted linear functions (Table I). (B) Mean durations, estimate of $\pm SE$, and fitted quadratic functions (Table I).

a subject worked with lists drawn from sets of nine words of each length that were changed from day to day; the sets were chosen so that a subject encountered the two members of any embedded pair on different days.

Each list was presented on three successive trials; because repetitions had little effect on either latency or duration, the data were averaged over repetitions. Each subject contributed about 12 observations per list length per word length per day.

Latency functions in this experiment (Figure 3A) were significantly nonlinear, but since this is the only experiment among many using lists of up to six words, in which we have observed such nonlinearity, we feel justified in describing the latency data in terms of parameters of the fitted linear functions. Slopes of these functions (in units of milliseconds per word) for one-and two-syllable words are almost identical, confirming the suggestion derived from the other two experiments; the difference is 0.9 ± 1.1 msec/word. However, the *mean* latency was influenced to a small but statistically significant extent by number of syllables: Mean latency for lists composed of two-syllable words was 4.5 ± 1.3 msec longer. (Experiments by other investigators have perhaps not been sensitive enough to detect differences as small as this; hence the earlier claim of no effect.) In an idealized description of these data one can take the slope difference to be zero and assert, therefore, that the effects of syllables per word and words per list are *additive:* Each factor has the same effect regardless of the level of the other.

The relation between latency functions for the lists containing one- and two-syllable words is stable with increasing practice, as shown by Figure 4. Both

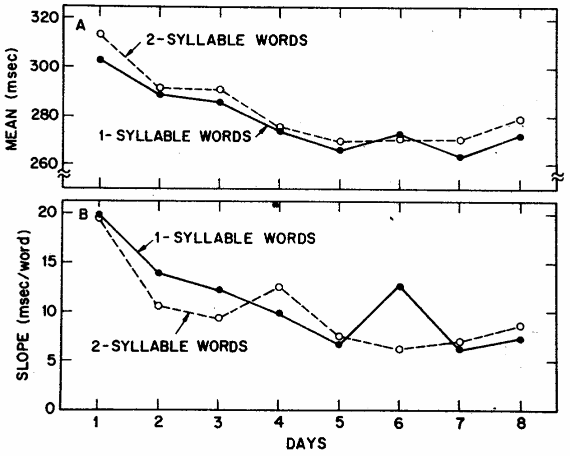

Figure 4 Effects of 8 days of practice on two aspects of latency (Figure 3A) for lists of length one to four of one- and two-syllable words. (A) Mean latency averaged over four list lengths, versus days; about 200 observations per point. (B) Slope of fitted linear latency function, versus days.

means and slopes decrease to what appear to be asymptotes, and, while there is no consistent ordering of slopes across days, the mean latency for lists of one-syllable words is smaller on 7 out of 8 days.

As one would expect, and despite its small effects on latency, number of syllables had a large effect on the duration function (Figure 3B). Data from both conditions are again well described by quadratic functions: The large duration difference is localized (see Table I) primarily in the constants (1-intercepts; 122.4 versus 173.4 msec) and linear coefficients (91.9 versus 180.4 msec/word); the quadratic coefficients (13.6 versus 12.0 msec/word^2, with a difference of 1.6 ± 0.9 msec/word^2) are almost identical, confirming the earlier findings.

D. Some Findings from Other Experiments

Here we list a few of our findings from other experiments that are especially pertinent to the interpretation of the latency effect.

1. Asymptote of the Latency Function

As the length of well-learned sequences is increased beyond about six words, the latency ceases to increase in an orderly manner and tends, instead, to fluctuate about an asymptotic value. (For novel sequences the range of the latency effect is of course limited in a different way: Once the immediate memory span is exceeded, both latency and error rate increase drastically.)

2. Insensitivity to an Additional Load on Short-Term Memory

In one experiment subjects memorized two lists on each trial: The first was to be recited fast, and the second was to be recited at a leisurely pace afterwards. The combined length of the two lists was never greater than five items. The presence and length of the second list had virtually no effect on the latency for reciting the first list (or on its duration). The latency therefore depends not on the load on short-term memory before or after the signal, but on the amount that must be said fast.

3. Invariance across Recite Signals of Different Modalities

In an experiment comparing visual and auditory recite signals, the difference between slopes of the resulting latency functions for lists of one to five words was negligible and nonsignificant.

4. Invariance across Changes in Time Uncertainty

In the experiments described here we attempted to minimize subjects' time uncertainty about when the recite signal might occur by using fixed foreperiods and countdown signals. In an experiment in which this procedure was compared to one with a highly variable foreperiod, we found that although the increase in time uncertainty was effective, in the sense that it caused an increase in the mean of the latency function, the change in its slope was negligible and nonsignificant.

5. *Effect of Interpolating Words without Primary Stress*

We have studied the effects of inserting unstressed connectives (*and, of, or*) between the successive nouns in a list, and of inserting words with little stress (*by, and, is a, minus*) between the successive (stressed) digits in a list. Neither of these operations increased the slope of the latency function (in milliseconds per noun or per digit) above its normal value; however, they did increase the mean latency. Thus, whether a list is augmented by extra unstressed syllables, or by separate words that carry little or no stress, the result is the same: There is a small increase in the mean, but no effect on the slope.

III. Hypotheses about the Latency Effect

The findings presented thus far provide us with two principal facts to be explained. First, the latency for rapid reciting of a prespecified utterance increases approximately linearly and at a rate of about 10 msec/word (or primary stress), over a range of one to five words. Second, over the same range, the more words in the utterance, the slower the rate at which it is recited. We begin with a discussion of some of the alternative hypotheses that we have considered for the latency effect and that have guided our experiments.

A. Time-Sharing: Readiness versus Rehearsal

The first two hypotheses reflect the possibility that the latency effect results from the load imposed by lists of increasing length on a short-term memory store, as traditionally conceived.[9] Suppose that while the subject awaits the signal he divides his time between maintaining his memory of the list (by covert rehearsal, for example) and being ready to respond to the signal. If the reaction signal occurs when he is in the *maintenance state* then it takes time to shift into the *ready state*. Increasing the list length may increase either the proportion of time spent in the maintenance state or the time taken to switch out of it. Given this hypothesis, and assuming that the subject has conrol over which state he is in, one would expect that increasing the time uncertainty associated with the reaction signal would make it more likely that the subject was in the maintenance state

---

[9] Here and elsewhere (where we use the term "ordinary short-term memory") we refer to the "short-term store" of Atkinson and Shiffrin (1968), or "primary memory" (see, e.g., Craik & Levy, 1976; Crowder, 1976).

when the signal occurred and would therefore increase the size of the latency effect (indexed by the slope of the latency function). The invariance of the slope, found when time uncertainty was manipulated experimentally, argues against this hypothesis. (In addition, the subjects asserted that they were not consciously rehearsing as the time approached when the signal might be presented.)

B. Capacity-Sharing and the Load on Short-Term Memory

A related possibility is that the longer the list, the more of a limited "capacity" (rather than time) has to be devoted to maintaining it in memory, and this reduces the capacity available for either processing the recite signal or having the first item ready. Several of our findings argue against this hypothesis (as well as the previous one). First, the similarity of latency functions for weekdays in normal, random, and repeating sequences shows that the capacity required to retain the list in memory is unimportant. Second, Baddeley, Thomson, and Buchanan (1975) have shown that the limit on the number of unrelated words that can be retained in short-term memory can be described in terms of the time it takes to say those words, so that brief words require less of the memory capacity than long words. Yet we have shown that the slope of the latency function is independent of word duration. Finally, by using an auxiliary memory load, we showed that increasing the number of words to be maintained in short-term memory, without changing the number to be recited fast, does not influence the latency.

C. Competition among Distinct Response Elements

According to the element-competition hypothesis, adequate preparation for the reaction signal requires that all the distinct elements from which the response is to be constructed be mentally "primed" or "activated" before the response begins (Lashley, 1951; Wickelgren, 1969). Suppose that all the primed elements then compete with any element that has to be produced, and that competition increases latency. The first element in a longer list has more competitors and hence a longer latency. Such a mechanism is particularly appealing because it can also readily explain the finding that the average rate at which the elements in longer lists are produced is lower. This hypothesis can be rejected on the basis of the typical latency effect found in the repeated weekdays condition, where list length was increased not by enlarging the set of *types* (and hence the number of *distinct* elements to be primed), but by adding to the number of *tokens* of the same type. Insofar as the primed elements are smaller than words, the absence of an effect of syllables per word on the latency-function slope can also be taken as negative evidence.

D. Information Transmitted

A fourth hypothesis is based on the observation that the amount of selective information in the response, in the sense of Shannon (see Garner, 1962), increases with the length of the list. Suppose that even though the list is specified in advance, the subject must resolve response uncertainty in selecting what to say at the time of response intiation. According to one version of this hypothesis (*sequence uncertainty*), the list must be selected from the set of all lists of the same length drawn from the same vocabulary. According to another version (*first-item uncertainty*), the first word must be selected from among all the words in the list. Under either version, if latency increased with the amount of information transmitted by the response, one would expect that for random lists, latency would increase with length. However, since an increase in the number of *repeated* words should not cause a corresponding increase in the amount of information, results from the weekdays experiment argue against this hypothesis.

E. Sequence Preparation (Version 1): Motor-Program Construction or Activation

According to the final and, we think, the most acceptable pair of hypotheses among those we consider, a representation of the entire response appropriate for controlling its execution (a *program*) is constructed before the response starts. The program consists of a set of linked *subprograms,* one for each *unit* of the response. The measure of the length of the program is the number of subprograms it contains. We suppose that the program is retained in a special *motor-program buffer* that is distinct from ordinary short-term memory. This memory state (or code or structure) is not sensitive to factors such as familiarity of the response, similarity or identity among its elements, its duration as such, or the extent to which short-term memory is otherwise occupied.

The first version of the sequence-preparation hypothesis is in the spirit of the "memory drum" model proposed by Henry and Rogers (1960): Either part or all of a process by which the program is constructed, or else a process of activating a previously constructed program (by "loading" it, for example), is not (and perhaps cannot be) started before the signal. Furthermore, this process is completed before the response begins and has a duration that increases linearly with the number of units in the list. Either the duration of the process is independent of subprogram length or subprogram length is independent of unit size; otherwise we would have seen an effect on the latency slope of syllables per word.

Why should a process of program construction or activation not be completed *before* the signal, when the subject knows in advance exactly what has to be said? We offer two speculative reasons:

1. Constructing or activating a motor program might be inherently tied to its execution: Once the program is ready to be used, execution follows automatically and is hard to inhibit. If the preparation process took place before the signal, the subject would then respond on catch trials. To avoid this, preparation must await the signal. If program activation involved loading the program into a delay-line memory, for example, and execution occurred when the information emerged at the other end of the delay line, the system would have the required property that activation causes execution.

2. The contents of the motor buffer might be subject to rapid decay, in which case the program would have to be set up immediately before use. (Furthermore, processing of the reaction signal might interfere with maintenance of information in the motor buffer.)[10]

F. Sequence Preparation (Version 2): Subprogram Retrieval

According to the second version of the sequence-preparation hypothesis, the construction and any necessary activation of the program as a whole are accomplished before the signal. However, only after the signal occurs is the subprogram for the first unit retrieved or located in the program. Mean retrieval time increases linearly with the number of subprograms contained in the buffer. This could arise because the retrieval is accomplished by means of a sequential search (analogous to mental search processes that have been proposed for other domains; e.g., Sternberg, 1969b) through the set of subprograms or through a directory of subprogram "addresses." Alternatively, retrieval time could increase with the number of subprograms because capacity has to be shared between a parallel search or a direct access process, and maintenance of the motor program; the property of linearity is less readily associated with such mechanisms, however. Possible reasons why retrieval of the first subprogram might have to await the signal are similar to the two reasons given above.

The retrieval version of the sequence-preparation hypothesis has a feature not shared with Version I: It not only explains the latency effect, but also leads naturally to an account of the slower production rate for longer lists, if the same retrieval mechanism is assumed to apply to each of the other units in the list as to the first. Insofar as the relations between latency and duration data are consistent (inconsistent) with such an account, they support the retrieval (activation) version.

---

[10] Neurophysiologists have argued that parts of the nervous system that may be implicated in the sequencing and timing of rapid movements can retain information only briefly. Also, certain neural structures, such as the parallel fibers in the cerebellum, appear to function as delay lines—though the delays they generate may be too short for present purposes (see Kornhuber, 1974; Eccles, 1969).

Suppose, for example, that the mechanism responsible for the length effect is a simple search process.[11] Then there are two obvious ways in which latency and duration might be related. If the same set of subprograms is searched, regardless of how many units have already been produced, then we expect that the mean time between the production of one unit and the next will increase linearly and at the same rate with list length as the latency does. On the other hand, if the contents of the buffer shrink as the response proceeds, then it is easy to show that the mean time between one unit and the next (averaged over the list) will again increase linearly, but at half the rate at which the latency increases. (These statements are valid without further assumptions if the search process is *exhaustive* [Sternberg, 1969b]; if it is *self-terminating* they depend on the order of the search meeting certain requirements.) A similar argument can be made if the mechanism is based on capacity sharing; in that case, we have a variety of element competition (see Section III,C) that applies to identical as well as distinct elements. We shall consider the duration data below, in light of these issues.

IV. Elaboration of the Sequence-Preparation Hypotheses

A. Nature of the Programming Unit in Speech

Both versions of the sequence-preparation hypothesis incorporate the idea that the response latency increases by some amount for each part (subprogram) of the motor program, where each subprogram controls one of the *units* in the response.[12] What is the nature of these response units? (They need not correspond to what we have called *elements*—convenient but arbitrary response segments.)

---

[11] The idea that in proceeding through a set of linked subprograms a search is used (rather than a process involving direct access by one subprogram to the next) would be superfluous if subprograms were stored in the same order in which they were to be used. An example of a theory of the control of serial order that is more congenial to the search idea is Wickelgren's (1969) context-sensitive associative theory of speech production, whose units could be stored in a random order and still support an ordered serial response, because each unit is stored with tags that indicate the preceding and following units.

[12] An alternative possibility for the construction version of the hypothesis is that the latency depends on the length of a higher-level program from which the motor program is constructed. Consider the following analogy: There is a "source program" (e.g., a series of coded representations in short-term memory) and an "object program" (the motor program) that is compiled from it and held in the motor buffer. Compiling requires fetching source units, translating them, and sending the resulting object units. Object units need not correspond one to one with source units. The rate of compiling could primarily reflect the rate of fetching, the rate of translating, or the rate of sending. In the construction version of the hypothesis, we cannot assign the latency effect unambiguously to one of these. Thus, the latency might, for example, depend on the number of words (a possible source unit) or the number of stress groups (a possible object unit). For the activation or retrieval versions of the hypothesis, however, the latency effect should reflect object units.

Recall that variation in the number of syllables per word had no effect on the latency increment per word. Since this manipulation indirectly alters the duration of a word, the number of articulatory gestures it contains, and the number of syllables, the unit cannot be a *speech segment of specified duration,* the *articulatory gesture,* or the *syllable.* Two of the remaining alternatives are the *word,* and the *stress group* or "metric foot" (a segment of speech associated with a primary stress). We have seen that the increase in latency with list length depends on the number of words in the list with primary stress and not on whether other words, with little or no stress, are inserted. Our present conclusion is that each subprogram controls a stress group, and we shall tentatively assume this in the following sections. The fact that the unit that underlies performance in our experiments appears to be articulatory rather than semantic seems to us to support further the interpretation of the latency effect in terms of a motor program.

B. A Basis for the Syllables Effect

We have seen (Section II, C and D) that although there was no effect on the latency-function *slope* (in milliseconds per stress group) of increasing the number of syllables per word or of inserting unstressed words, these operations did add an approximately constant increment to the latency for lists of each length, thereby increasing the latency-function *mean.* How can this increase be explained in the context of the sequence-preparation hypotheses? One possibility is that while activation (of the entire program) or retrieval (of the first subprogram) is completed in a time determined by the number of units, each unit has to be further *unpacked* into its constituents (syllables or articulatory gestures, for example) before it can be executed. Unpacking can be regarded either as advance planning or as retrieval, at a lower level of the response hierarchy.

Suppose that duration of the unpacking process increases with the *size* (number of constituents or duration) of a unit. Since only the first unit must be unpacked before the utterance begins, the unpacking operation prolongs the latency by the same amount for all list lengths. The latency function therefore increases in mean, but not in slope, when we increase the unit size. One implication that could be used to test the unpacking idea is that for a list of specified length whose units are of mixed size, the latency should depend only on the size of the first unit.

V. Analysis of the Duration Function

A. Quantitative Representations for Duration Data

Our second principal finding is that the duration function accelerates with sequence length, which suggests that the average interval between the starting times of successive elements is greater for longer sequences; this property also characterizes typewriting (see Section VI). We have seen in Section III,F that the quantitative characterization of the duration effect may help to select among alternative hypotheses for the latency effect. In the present section we introduce some concepts and notation that will facilitate the analysis and comparison of duration data in speech and typing.

The existence of the duration effect suggests that it might be useful to examine the individual time intervals between successive elements that add together to generate the observed duration. Such examination, however, requires measuring the interval between single time points associated in an invariant manner with individual elements (a *measurement-invariance requirement*); this is difficult in experiments involving rapid speech, but easy for typewriting, where the exact time of each key depression can be readily determined.[13] Let these times in a sequence of length n be $T_{1n}, T_{2n}, \ldots, T_{nn}$, with the reaction signal specifying the time origin so that T_{1n} is the latency, L_n. (Note that if we regard each response element as *ending* with the key depression, then the measured latency in typing incorporates the duration of the first element, unlike the latency measure in speech, which does not incorporate the duration of the first word.) The $n-1$ time intervals between successive elements are then $T_{2n}-T_{1n}, T_{3n}-T_{2n}, \ldots, T_{nn}-T_{n-1,n}$, which we denote $R_{2n}, R_{3n}, \ldots, R_{nn}$, respectively. A useful measure of production rate that can be estimated from the duration, $D_n = T_{nn} - T_{1n}$ of an entire sequence of length n, and which is easy to relate to the $n-1$ time intervals between successive elements, is the *mean* such time interval,

$$R_{\cdot n} = \frac{1}{n-1}\sum_{k=2}^{n} R_{kn} = \frac{1}{n-1}\sum_{k=2}^{n} (T_{kn} - T_{k-1,n}) = \frac{1}{n-1}(T_{nn} - T_{1n}) = \frac{1}{n-1} D_n, \quad (n \geq 2). \quad (1)$$

We use "*R*" because $R_{\cdot n}$ and R_{kn} are measures of *rate;* note, however, that they denote time per response element, not elements per unit time.

The rate function $R_{\cdot n}$ provides an alternative representation of duration data that has some useful properties. First, if D_n increases as a quadratic function of length, then, as we shall see below, $R_{\cdot n}$ increases linearly.[14] This simplicity of form, together with the fact that $R_{\cdot n}$ has a smaller range of variation than D_n, makes any systematic deviations from the fitted function more apparent and facilitates comparisons between functions. In addition, in those instances where

---

[13] As a starting point, one can decompose the movement sequence in typing a list of letters into a series of elements, each ending at the moment a key depression is detected. Then, by definition, the single measured time point always marks the end of a movement element, and so the requirement of an invariant relation between time point and element is satisfied. It is possible, however, that this initial decomposition of the movement sequence is not the best one theoretically, even though it facilitates measurement; one test is whether the data are orderly and easily interpretable. We believe that for the movement sequences in rapid speech the appropriateness of any particular decomposition is even less obvious.

[14] An alternative measure with this property but less desirable in other respects is the first difference function, $D_n - D_{n-1}$, which is similar to the derivative.

the duration variance increases most dramatically with n, we have observed the $R_{\cdot n}$ variances to be more homogeneous.

Estimation of $R_{\cdot n}$ as defined in Eq. (1) is straightforward for typing data, but to apply that definition to our speech experiments the duration measures first need correction. Let T_{kn}^* and T_{kn} represent, respectively, the starting and ending times of the kth word in a response of length n. For the response as a whole the measured starting time T_{1n}^* (appropriate for a measure of latency) and the measured ending time T_{nn} do not mark corresponding points in the first and last response elements; these measures therefore fail to meet the measurement-invariance requirement mentioned previously in this section. Let us identify the *end* of each word as the desired (invariant) time point. Then the measured duration $D_n^* = T_{nn} - T_{1n}^*$ includes not only the sum $D_n = T_{nn} - T_{1n}$ of the $n - 1$ intervals from the end of one word to the end of the next, but also the time $T_{1n} - T_{1n}^*$ from the start to the end of the first word. To estimate D_n from D_n^* we must therefore subtract an estimate of this extra time. If we assume that the duration of the first word is independent of n, then $D_1^* = T_{11} - T_{11}^*$ provides the desired estimate, so that $D_n = D_n^* - D_1^*$, and we have

$$R_{\cdot n} = \frac{1}{n-1} D_n = \frac{1}{n-1} (D_n^* - D_1^*), (n \geqslant 2). \qquad (2)$$

If D_n is actually quadratic, but the duration of the first word depends on n, then it is likely that D_n^* would differ systematically from a fitted quadratic function and that $R_{\cdot n}$ would be systematically nonlinear. (A dependence of the duration of the first word on n, particularly on $n = 1$ versus $n > 1$, could come about from failure of the element-invariance requirement discussed in Section I,A.) Note that Eq. (2) can be regarded as a generalization of Eq. (1); in a case such as typewriting, $D_1^* = 0$, so $D_n = D_n^*$.

Now let us consider a quadratic duration function as fitted to the data in Figures 1B, 2B, and 3B:

$$D_n^* = \alpha + \beta(n-1) + \gamma(n-1)^2, (n \geqslant 1). \qquad (3)$$

We have written D_n^* as a function of $n - 1$ rather than of n because the parameter α then represents D_1^* (which is zero for the case of typewriting) and because the rate and duration functions are then related in a simple way; from Eqs. (2) and (3) we get

$$R_{\cdot n} = \beta + \gamma(n-1), (n \geqslant 2). \qquad (4)$$

Thus, a quadratic duration function implies a linear rate function: The quadratic coefficient, γ, in D_n^* represents the amount by which the average interval between one element and the next increases for each element added to the response, whereas the linear coefficient, β, represents a "base" value of the average inter-element time, to which the successive increments are added. If the duration function were linear, we would have $\gamma = 0$, and the rate function would then be a constant.

B. Analysis of Durations in Experiments on Speech

Latency and duration functions that were fitted to the data from the three speech experiments are summarized in the top four lines of Table I. As already noted, the quadratic coefficients in the four duration functions are remarkably similar in magnitude, despite substantial differences in word duration. This implies that the rate at which the mean time between successive words increases with sequence length is the same for words containing different numbers of syllables. To reveal this more clearly, we have displayed the observed and fitted rate functions $R_{\cdot n}$ for our third experiment (see Figure 6A). When we describe the production rate as a function of the number of words (rather than, for example, the number of syllables) the two rate functions are almost perfectly parallel, just as for the latency functions (see Figure 3A), despite their considerable difference in mean: The slope difference between the fitted functions is

Table I
Fitted Latency and Duration Functions from Four Experiments [a]

Experiment or condition	Latency function, $L(n) = \eta + \theta(n-1)$	Duration function, $D(n) = \alpha + \beta(n-1) + \gamma(n-1)^2$
Speech		
Ascending numbers	$263.3 + 12.6(n-1)$	$101.1 + 57.6(n-1) + 12.2(n-1)^2$
Weekdays (mean)	$259.9 + 8.8(n-1)$	$168.8 + 183.1(n-1) + 12.6(n-1)^2$
Monosyllabic nouns	$261.5 + 11.1(n-1)$	$122.4 + 91.9(n-1) + 13.6(n-1)^2$
Disyllabic nouns	$267.4 + 10.1(n-1)$	$173.4 + 180.4(n-1) + 12.0(n-1)^2$
Typewriting		
Alternating hands	$229.7 + 14.9(n-2)$	$71.9(n-1) + 15.2(n-1)^2$
One hand	$231.2 + 4.1(n-2)$	$142.9(n-1) + 14.1(n-1)^2$

[a] Values are in milliseconds. The element count, n, represents the number of words (speech experiments) or the number of letters (typing experiment) in the response. Parameters θ and η of $L(n)$ were fitted by least squares to data for $n > 1$ (speech) or for $n > 2$ (typing). Estimates of β and γ were determined by least-squares fitting of a line to the rate function $R_{\cdot n}$. For speech, the constant α was then determined by least-squares fitting of a quadratic function with specified β and γ; for typing, α was set at zero, corresponding to the measured duration of a single response.

only 1.6 ± 0.9 msec/word.[15] That is, the effects on the mean time between successive words of the number of syllables per word and the number of words in the response are almost perfectly additive: The first term, β, depends only on word length, whereas the second term, $\gamma(n-1)$, depends only on number of words. The simplicity of our results, when described in this way, supports the view that it is the number of words or stress groups in the response, not the number of syllables or articulatory gestures, that determines the decline in production rate with response length—the same response unit that we have seen (in Section IV,A) to be implicated in the growth of latency.

The results in Table I also show that the rate and latency effects—measured by the parameters γ and θ, respectively—are remarkably close in magnitude. This is consistent with the proposal that the two effects are actually the same and are generated by a common mechanism, as is assumed in the subprogram-retrieval version of the sequence-preparation hypothesis, with a nonshrinking buffer (Section III, F). However, although similar in many ways, our results from measurements of typewriting will force us to question the generality of identical effects.

VI. An Experiment on Typewriting

For several reasons we decided to study the sequences of rapid finger movements in typewriting using essentially the same paradigm as in the speech experiments. First, typing readily permits measurement of the individual time intervals between successive response elements, in addition to overall response duration. Second, although potential artifacts due to measurement error may never be entirely eliminated, they are at least likely to be different in typing than in speech. In speech for example, measurement delays might be influenced by loudness, sequence length might influence the volume of air in the lungs, and it is difficult to apply an objective criterion of response accuracy. These problems are obviously not critical in typing, but others might be, such as variations in the starting position of the hands or in the movements used to press a particular key. Third, we wished to see to what extent our findings generalized to a very different performance with a different training history. It can be argued that the production of a spoken word is much more complex than the pressing of a key, because it requires the exquisitely precise timing and coordination of a large

---

[15] Both sets of data are slightly concave downward and, although we have not found this to be true in all our speech experiments, we have observed this effect often enough to make us suspect that it is real. Examination of the first differences, $D_n - D_{n-1}$, makes us suspect that the nonlinearity of $R._n$ results from a violation of the element-invariance requirement (see Section I,A), which causes the estimation error discussed in Section V,A: The measured duration, D_1^*, of a single isolated word may be somewhat longer than the duration of the first word in a list that contains more than one word.

number of diverse muscle systems that control respiratory, laryngeal, and multiple articulatory mechanisms (see, e.g., Kent & Moll, 1975; MacNeilage & Ladefoged, 1976). Furthermore, while normal speech rates are far slower than those obtained in our experiment, typists are trained to achieve high rates outside the laboratory.

A. Procedure

The procedure was similar to that used in the speech experiments. On each trial a row of from one to five different letters was first displayed for 1.0 sec. The display was followed by a fixed delay of 2.4 sec. On 80% of the trials a brief tone burst occurred at the end of the delay, signaling the subject to start typing the letter list. This reaction signal was preceded at 0.7-sec intervals by two brief noise bursts, which served as "countdown" signals. Keypress responses on the remaining 20% of the trials, on which the reaction signal was omitted, were regarded as errors. Again, the procedure was designed to encourage subjects to complete their responses as soon as possible after the signal while maintaining a low error rate. (The mean percentage of trials on which errors occurred was 2.3%; these few errors will not be considered further.) The time recorded for the depression of a key was determined by when it was detected by an electronic keyboard.[16]

Each letter list was presented on three successive trials; because repetitions had little effect on either latency or duration, the data were averaged over repetitions. Lengths and compositions of the lists were varied from one group of three trials to the next by means of a different balanced randomization for each subject.

Since Lahy's (1924) pioneering "Motion Study in Typewriting" and similar work reported by Coover (1923), it has been known that letter bigrams that are typed by fingers on alternate hands can be produced at faster rates than bigrams typed by fingers of the same hand. To explore responses containing elements whose durations differed, as we did in speech by manipulating the number of syllables per word, we used pure one-hand sequences in some blocks of trials and pure alternating-hand sequences in others; in both conditions, left and right hands were used equally often. All sequences were drawn from the same 16 letters, and average bigram frequencies in English were equated in the two conditions; we used only bigrams that actually occur in English.

The subjects were four female professional typists employed at Bell Laboratories, with test rates in prose typing of about 90 words per minute (or about 7.5 strokes per second). The data to be presented were obtained on 2 days of testing, after 1 day of practice with the same types of material. During the 2

---

[16] Our keyboard had "N-key rollover," which permits it to detect a keypress with a neglible delay after it occurs, regardless of whether other keys remain depressed (see Kallage, 1972).

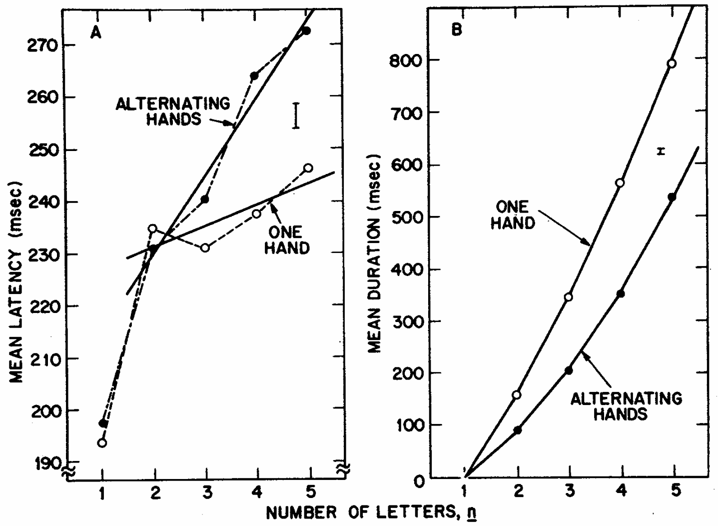

Figure 5 Typewriting of lists of letters typed by fingers on alternating hands and by fingers on one hand only: The results are averaged over four subjects; about 200 observations per point. (A) Mean latencies, estimate of ±*SE*, and linear functions fitted to data for $n \geqslant 2$ (Table I). (B) Mean durations, estimate of ±*SE*, and fitted quadratic functions (Table I).

days, each subject provided about 50 observations for each of the five list lengths per condition.

B. Latencies of Typing Responses

Figure 5A and Table I show that response latency again increased with sequence length, but in a different manner in the two conditions. The nonlinearity (latency for $n = 1$ falls below a fitted line) that is present in both sets of data is significant for the one-hand condition only. We have chosen to fit linear functions to the data for $2 \leqslant n \leqslant 5$ in both cases, however, to achieve comparability and because the standard error of the slope for the alternating-hand data, based on between-subject differences, is reduced by a factor of 2 when L_1 is omitted.[17] To emphasize the omission of L_1, in Table I we represent the fitted lines as functions of $n - 2$ instead of $n - 1$; the constant terms then represent intercepts at $n = 2$. Despite the irregularity of the function for the one-hand condition, estimated slopes are very similar across subjects: The mean and standard error are 4.1 ± 1.3 msec/letter for this condition (significantly greater than zero) and 14.9 ± 1.7

---

[17] In a replication of this experiment with different subjects we obtained very similar results.

msec/letter for the other, with a difference of 10.8 ± 1.8 msec/letter that is highly significant.

Why should either of our latency functions show a discontinuity between a single keystroke and multiple keystrokes? And why should such an effect occur more strongly in typing than in speech? One possibility is that as sequence length is changed from $n = 1$ to $n = 2$, the response elements fail to satisfy the element-invariance requirement (see Section I,A). In particular, a single keystroke may not be equivalent to either a beginning or terminating keystroke, especially in the one-hand condition. One possible reason for the contrast with our speech data is that while the production of even a single monosyllabic word typically involves multiple articulatory gestures and therefore requires precise control over timing and coordination, this may be true only of sequences of two or more keystrokes, because of the relative simplicity of a single keystroke.

There is no reason to believe that the letter strings in the two conditions place systematically different loads on ordinary short-term memory; the difference between the latency functions therefore provides further evidence against the time-sharing and capacity-sharing hypotheses discussed in Section III.

C. Response Durations and Interstroke Intervals

Both duration functions, shown in Figure 5B, are significantly nonlinear but are well described by the fitted quadratic functions, in remarkable agreement with the speech data. Our subjects responded well to the exhortation that they complete their responses quickly: In the alternating-hands condition they produced about 9.3 strokes per second—a higher rate for these meaningless letter strings than they averaged in continuous prose. As expected, the one-hand condition produced substantially longer durations, but, as in speech, element duration appears not to affect the quadratic coefficients. The rate functions in Figure 6B again make the simplicity of the duration data more apparent and facilitate comparison to the speech data in Figure 6A. The fitted linear functions that describe both sets of typing data so well are separated by about 70 msec and are almost perfectly parallel, with a mean slope difference of only 1.0 ± 2.5 msec/letter. Thus, the effects on the mean time between successive strokes of the number of strokes and of the nature of the transition from one stroke to the next are almost perfectly additive. More experiments are needed before we can interpret the similarity of rate-function slopes in milliseconds per keystroke for typing and in milliseconds per stress group for speech.

For the alternating-hands condition, we observed still another property of the speech data that also characterizes typing: The sizes of the latency and rate effects, measured by the parameters θ and γ, respectively, are very similar, with $\hat{\gamma} - \hat{\theta} = 0.3 \pm 5.1$ msec/keystroke. However, the conjecture that the two effects

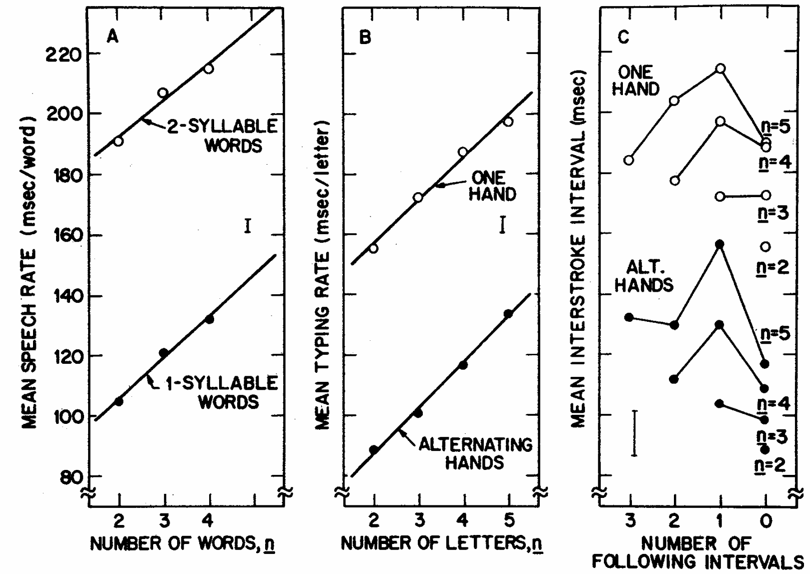

Figure 6 Time intervals between successive response elements in speech and typing. (A) Rate functions (mean interval versus response length) for speech and estimate of $\pm SE$. Fitted lines: $91.9 + 13.6(n-1)$ for one-syllable words and $180.4 + 12.0(n-1)$ for two-syllable words. The data are from Figure 3B. (B) Rate functions for typing and estimate of $\pm SE$. Fitted lines: $71.9 + 15.2(n-1)$ for letter lists typed with alternating hands and $142.9 + 14.1(n-1)$ for letter lists typed with one hand only. The data are from Figure 5B. (C) Mean interstroke interval for each serial position in letter lists of lengths $n = 2, 3, 4$, and 5 typed with alternating hands and with one hand only, and estimate of $\pm SE$. The results are averaged over four subjects; about 200 observations per point. Intervals early in a list have more intervals following them and appear toward the left of the plot. The *SE* estimate is based on the *serial-positions* × *subjects* interaction. The points in B represent the means of the functions in C.

are in general mediated by the same mechanism, supported by these data as well as by the results from the speech experiments, is called into question by the one-hand data, where we observe a significant difference of $\hat{\gamma} - \hat{\theta} = 10.1 \pm 3.0$ msec/keystroke between these measures of the rate and latency effects. More work is needed to decide whether to give up the conjecture and explain the similarity of γ and θ values in some other way or to seek a special explanation for the dissociation of effects found only in one typing condition.

Figure 6C shows, for responses of length $n = 2, 3, 4$, and 5 in each condition, the individual intervals, $R_{kn} = T_{kn} - T_{k-1,n}$, between successive keystrokes that contribute to the means, $R_{\cdot n}$, of Figure 6B. The means of each of the $n - 1$ successive interstroke intervals contained in responses of length n are shown as functions of the number of *following* intervals—a number that appears to be a more powerful determinant of the forms of these functions than the number of preceding intervals.[18]

In every case, intervals followed by the same number (0, 1, or 2) of other intervals are longer when they are contained in longer responses, demonstrating the strength of the effect of sequence length on stroke rate. Note that even the time from the first stroke to the second (not lined up in Figure 6C) is systematically greater for longer sequences. For sequences of four and five letters, the stroke rate is slower toward the middle of the sequence than at either end.

The patterns of interstroke intervals are very similar in the two conditions. To a first approximation, then, the two sets of R_{kn} functions differ by just a constant: The effect on interstroke interval of whether interstroke transition is within or between hands is roughly additive with the effect of serial position, at the same time as the effect on *mean* interstroke interval of transition type is almost perfectly additive with the effect of length. One way to explain such instances of additivity is to assume that the effects on interstroke interval of length and serial position, in the first place, and of transition type, in the second, are mediated by different *stages* of processing—separate operations that occur sequentially (Sternberg, 1969a).

An important challenge for any theoretical account of these duration data is to explain jointly the simplicity of performance at the macroscopic level (linearity with length at the level of the means) and the greater complexity at a microscopic level (nonlinearity with serial position).

VII. Summary of Findings and a Tentative Model for the Latency and Duration of Rapid Movement Sequences

A. Summary of Method and Findings

We begin this section by summarizing the main facts about rapid speech and typewriting that our experiments thus far have revealed. Subjects made *responses* composed of short sequences of equivalent *response elements,* either spoken words or keystrokes. They were rewarded for completing these responses as quickly as possible after a reaction signal. We varied the *response length* (number of elements) over a range from $n = 1$ to $n = 5$. We also manipulated *element size* (or duration) by varying either the number of *constituents* (syllables per word) or the nature of the required movements (successive keystrokes by same versus alternating hands). The two main measures were *response latency,* L_n, from the reaction signal to the first element and *response duration,* D_n, from the first element to the last. We found that a useful representation of duration data is provided by a *rate function,* $R_{\cdot n}$, that describes the mean time interval between one

---

[18] This might be taken as further evidence of the influence of advance planning.

element and the next, averaged over the $n - 1$ intervals in a response. In typing we also measured the individual time intervals, R_{kn}, between successive elements as a function of serial position, k, within the response.

We shall tentatively assume the correctness of our conclusion that in speech the *stress group* (a segment of speech associated with one primary stress) is the theoretically relevant response *unit*, based on the observation that it is in terms of this unit that our effects on both latencies and durations are most simply described. In typing we shall identify the single keystroke as the response unit.[19] For purposes of this summary, we shall also treat two of our findings as anomalies that, for the present, we shall not seek to explain. One is the shape of the latency function for one-handed typing, which is nonlinear over the $1 \leqslant n \leqslant 5$ range and shallow over the $2 \leqslant n \leqslant 5$ range. The other is the downward concavity of the latency functions in the third speech experiment. Given that we set these exceptions aside, the main facts that we need to explain (for lists of up to at least five units) are as follows:

1. L_n (mean latency) increases with n (response length).
2. The increase is approximately linear.
3. L_n increases with number of constituents per unit.
4. The effects on L_n of number of units and number of constituents per unit are additive (do not interact).
5. $R_{\cdot n}$ (mean time interval from one unit to the next) increases with n.
6. The increase is approximately linear.
7. The effects on $R_{\cdot n}$ of number of units and unit size are additive.
8. The rates at which L_n and $R_{\cdot n}$ increase with n are similar.
9. R_{kn} (mean time interval from unit $k - 1$ to unit k in a response of length n) changes nonmonotonically with serial position, k.
10. The effects on R_{kn} of serial position and unit size are approximately additive.

B. A Tentative Model

Based on qualitative aspects of our data, we found the sequence-preparation hypothesis (Sections III,E and III,F) to be the only survivor among the mechanisms we considered for the latency effect. This observation is the starting point for our tentative model. Next, because we are taking the similarity of the slopes of L_n and $R_{\cdot n}$ to be more than a coincidence, we prefer that the model account for both latencies and rates by means of the same mechanism. (If a dissociation

---

[19] Note that for speech we happened to define an *element* (a response segment of arbitrary size) in such a way that it corresponded to what our results later suggested was a *unit* (a theoretically relevant response segment). Thus, in the speech experiments of Sections II,A–II,C our element was a *word* (rather than, for example, a syllable or a segment of specified duration), which was equivalent to a *stress group* in our materials. The convenient result is that response length, n, denotes the number of units as well as the number of elements, and unit size (in number of constituents or duration) is equivalent to element size.

between these slope parameters were discovered in later experiments, our preference might change, of course.) This decision, in turn, favors the second version of the preparation hypothesis (subprogram retrieval) over the first (program construction or activation) because Version 1 would require that the entire program be reconstructed or reactivated before the execution of each unit—a requirement that appears implausible to us. Our choice of model for the latency effect therefore depends on our findings about durations.

The particular retrieval mechanism suggested by our results is *self-terminating* sequential *search* through a *nonshrinking* buffer, rather than, for example, a process of direct access whose speed is limited by a capacity that must be shared among all the subprograms. *Search* seems to lead more naturally to linearity of time versus length, given the direct-access models of which we are currently aware. The buffer should be *nonshrinking* because otherwise $R_{\cdot n}$ would have half the slope of L_n, rather than approximately the same slope. The search is presumably necessary because subprograms are not arranged in the buffer in the order in which they must be executed. By assuming a *self-terminating* search rather than an exhaustive one (e.g., see Sternberg, 1969b), we are able to accomodate a wide variety of effects of serial position on R_{kn} (together with the approximate linearity and slope equality of L_n *and* $R_{\cdot n}$) by suitable assumptions about search order.

One aspect of search order is its variation from one trial to the next, which might depend, for example, on the order in which subprograms are stored in the buffer. Another aspect is its variation from the search for one subprogram to the next within a response: The search might start at the same location for each subprogram or at the location of the last subprogram retrieved, for example. In any case, it is the mean position of a subprogram in the search order when that subprogram is the one to be executed that determines serial-position effects. If the order is random, then all subprograms have the same mean position, and R_{kn} functions should be flat: The R_{kn} range, $\max\{R_{kn}\} - \min\{R_{kn}\}$, should be approximately zero. At the other extreme, if the order is fixed so that a subprogram in some one serial position is always found first and a subprogram in some other serial position is always found last, position effects are maximized and, ignoring sampling error, the R_{kn} range can be shown to be $2(n-1)\gamma$. The ratio of the observed range to its estimated maximum provides a measure of the magnitude of position effects. For the typing data shown in Figure 6C, we find the average value of this ratio for $n = 3$, 4, and 5 to be 0.04, 0.23, and 0.27, respectively, indicating that position effects were relatively small and implying that the order of the hypothesized search is closer to being random than fixed.[20] This finding is consistent with the approximate equality of L_n and $R_{\cdot n}$ slopes, which

---

[20] At present we can offer no convincing explanation for the shapes of the serial position functions. We can assert, however, that the observed shapes are not incompatible with a self-terminating search. Indeed, ignoring sampling error, it can be shown that if the R_{kn} range is no greater than $n\gamma$ then *any* serial-position function can be accommodated by a suitable self-terminating search order.

requires us to assume that the mean search position of the first subprogram is approximately equal to the mean search position of the others.[21]

The structure of our model is governed, in part, by application of the additive-factor method (Sternberg, 1969a) and is to some extent independent of details of the retrieval and unpacking processes. For the latency, the additivity of effects of length and number of constituents per unit suggests the existence of separate processing stages whose durations are additive components of the latency and which are influenced selectively by these factors. This leads us to propose a *retrieval stage* (influenced by the number of units, but not their sizes) followed by an *unpacking stage* (influenced by the number of constituents in a unit, but not by the number of units; see Section IV,B). Both of these stages follow operations that mediate detection of the reaction signal and the decision to respond.

Our assumption that execution of later units involves the same mechanisms as execution of the first unit leads us to propose that durations of the same retrieval and unpacking stages are included in R_{kn} and hence in $R_{\cdot n}$. That the effect of unit size on $R_{\cdot n}$ (which again is additive with the effect of number of units) is so much larger than its effect on latency is explained by the existence of a *command stage* (again not influenced by the number of units) during which the sequence of commands is issued that cause execution of the constituents of the response unit.

Exactly what the commands specify about the ensuing movements—whether muscle contractions or target positions, for example (MacNeilage & MacNeilage, 1973)—is not critical for the model. But we have to place constraints on the time relations between the command stage for a response unit, which we do not measure directly, and the execution of that unit. According to our model, successive command stages are not only sequential, but are temporally discrete, being separated by search and unpacking operations—operations whose average duration is estimated to be about 40 msec longer in 5-unit responses than in 2-unit responses. Yet the movements produced (and the resulting sounds, in speech) often appear to be smooth, continuous, and even overlapping (Kent & Minifie, 1977). To reconcile our stage model with this observation, let us assume that execution of unit k can be prolonged so as to continue in parallel with the search and unpacking stages associated with unit $k + 1$. We assume further that if it is prolonged, it is nonetheless interrupted or modified by the execution of unit $k + 1$, in such a way that it does not delay the completion of that unit. The execution of the final unit in the response may also outlast its command stage, but the amount by which it is prolonged is independent of response length, n. From these assumptions, it follows that the measured duration of a response differs

---

[21] One-hand typing might be an exception in this respect, where the subprogram for the first unit had a privileged position in the buffer, thereby producing a flatter latency function.

from the time between the beginning of the first command stage and the end of the last by at most a constant that is independent of n.

The model, then, introduces three types of processing stage. Let S_k denote the time to locate the subprogram for response unit k by means of a self-terminating search (or an alternative process, such as one involving direct access, that might later prove more appropriate). S_k depends on the number of units and on the serial position of unit k in the search order, but not on unit size. Let U_k denote the time between locating the subprogram and beginning the command sequence for unit k—the time to unpack the constituents—which depends only on the number of constituents per unit (length of the subprogram). Finally, let C_k denote the time to issue the sequence of commands that control unit k, which depends only on the size of the unit and not on either its serial position or the number of units.[22]

For the speech experiments, where the latency, L_n, is regarded as marking the *start* of the execution of the first response unit, we have

$$L_n = T^*_{1n} = T_b + S_1 + U_1, \tag{5}$$

where T_b denotes a "base time" during which the subject detects the signal and decides to respond. (Symbols are written in the same order as the corresponding stages are assumed to occur; see Section V,A for definitions of T_{kn} and T^*_{kn}.) For the typing experiment, where the latency is regarded as marking the *end* of the execution of the first response unit, we have[23]

$$L_n = T_{1n} = T_b + S_1 + U_1 + C_1. \tag{6}$$

For both kinds of experiment the inter–element time is

$$R_{kn} = T_{kn} - T_{k-1,n} = S_k + U_k + C_k, \; (2 \leqslant k \leqslant n). \tag{7}$$

This tentative model is consistent with all 10 of the facts listed above that our studies have revealed.

---

[22] An elaboration of these assumptions about C_k and/or U_k, for $k \geqslant 2$ is suggested by the weekdays experiment (Section II,B). There we found that responses consisting of repetitions of the same word differed only in duration from responses containing n different words and that this "fatigue" effect was limited to an increased value of the parameter β. Given our model, this finding would require that repetition have an effect on only the duration $U_k + C_k$ of unpacking and command stages of units after the first and that the increase in $U_k + C_k$ be as large on the first repetition as on later repetitions.

[23] The fact that latencies for the one-hand and alternating-hand conditions are ever similar (as they are for $n = 1$ and $n = 2$) may indicate that values of $U_1 + C_1$ are approximately equal in the two kinds of sequence. But the difference between the rate functions for the two conditions shows that even if this is the case, the values of $U_k + C_k$ ($k > 1$) must differ.

Acknowledgments

The research reported here could not have been done without the expert hardware and systems-software support provided by A. S. Coriell and W. J. Kropfl. We are also indebted to M. Y. Liberman for advice on phonetics, to J. B. Kruskal and D. E. Meyer for comments on the manuscript, and to the skilled typists from the Word-Processing Center and Secretarial Service at Bell Laboratories, Murray Hill, who served as subjects.

Reference Notes

1. Monsell, S., & Sternberg, S. The latency of short and rapid utterances: Evidence for response preprogramming. Paper presented at the Seventh International Symposium on Attention and Performance, Sénanque, France, August, 1976.
2. Rosenbaum, D. A. Human movement initiation: Selection of arm, direction, and extent. Unpublished manuscript, Bell Laboratories, 1978.

References

Adams, J. A. Issues for a closed-loop theory of motor learning. In G. E. Stelmach (Ed.), *Motor control: Issues and trends.* New York: Academic Press, 1976.

Anzola, G. P., Bertolini, G., Buchtel, H. A., & Rizzolatti, G. Spatial compatibility and anatomical factors in simple and choice reaction time. *Neuropsychologia,* 1977, *15,* 295–302.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press, 1968.

Baddeley, A. D., Thomson, N., & Buchanan, M. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior,* 1975, *14,* 575–589.

Berlyne, D. E. Conflict and choice time. *British Journal of Psychology,* 1957, *48,* 106–118.

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. A. Some variables influencing the rate of gain of information. *Journal of Experimental Psychology,* 1962, *63,* 105–110.

Butsch, R. L. C. Eye movements and the eye–hand span in typewriting. *Journal of Educational Psychology,* 1932, *23,* 104–121.

Callan, J., Klisz, D., & Parsons, O. A. Strength of auditory stimulus–response compatibility as a function of task complexity. *Journal of Experimental Psychology,* 1974, *102,* 1039–1045.

Coover, J. E. A method of teaching typewriting based on a psychological analysis of expert typing. *National Education Association Addresses and Proceedings,* 1923, *61,* 561–567.

Craik, F. I. M., & Levy, B. A. The concept of primary memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 4). Hillsdale, New Jersey: Erlbaum, 1976.

Crowder, R. G. *Principles of learning and memory.* Hillsdale, New Jersey: Erlbaum, 1976.

Duncan, J. Response selection rules in spatial choice reaction tasks. In S. Dornič (Ed.), *Attention and performance VI.* Hillsdale, New Jersey: Erlbaum, 1977.

Eccles, J. C. The coordination of information by the cerebellar cortex. In S. Locke (Ed.), *Modern neurology.* Boston: Little Brown, 1969.

Eriksen, C. W., Pollack, M. D., & Montague, W. E. Implicit speech: Mechanism in perceptual encoding? *Journal of Experimental Psychology,* 1970, *84,* 502–507.

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Greenwald, A. G., Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, 1970, *77*, 73–99.

Hayes, K. C., & Marteniuk, R. G. Dimensions of motor task complexity. In G. E. Stelmach (Ed.), *Motor control: Issues and trends*. New York: Academic Press, 1976.

Henry, F. M., & Rogers, E. E. Increased response latency for complicated movements and a "memory drum" theory of neuromotor reaction. *Research Quarterly of the American Association for Health, Physical Education and Recreation*, 1960, *31*, 448–458.

Hudgins, C. V., & Stetson, R. H. Relative speed of articulatory movements. *Archives Néerlandaises de Phonétique Expérimentale*, 1937, *13*, 85–94.

Kallage, R. Electronic keyboard design with N-key rollover. *Computer Design*, 1972 (February), 57–61.

Keele, S. W. Movement control in skilled motor performance. *Psychological Bulletin*, 1968, *70*, 387–403.

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

Kent, R. D., & Moll, K. L. Articulatory timing in selected consonant sequences. *Brain and Language*, 1975, *2*, 304–323.

Klapp, S. T. Implicit speech inferred from response latencies in same–different decisions. *Journal of Experimental Psychology*, 1971, *91*, 262–267.

Klapp, S. T. Short-term memory as a response preparation state. *Memory & Cognition*, 1976, *4*, 721–729.

Kornblum, S. Response competition and/or inhibition in two-choice reaction time. *Psychonomic Science*, 1965, *2*, 55–56.

Kornhuber, H. H. Cerebral cortex, cerebellum and basal ganglia: An introduction to their motor functions. In F. O. Schmitt & F. G. Worden (Eds.), *The neurosciences: Third study program*. Cambridge: MIT Press, 1974.

Lahy, J. M. Motion study in typewriting. In *Studies and reports* Series J (Educational) No. 3. Geneva: International Labour Office, 1924.

Landauer, T. K. Rate of implicit speech. *Perceptual and Motor Skills*, 1962, *15*, 646.

Lashley, K. S. The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley, 1951.

MacNeilage, P., & Ladefoged, P. The production of speech and language. In C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 7). New York: Academic Press, 1976.

MacNeilage, P. F., & MacNeilage, L. A. Central processes controlling speech production during sleep and waking. In F. J. McGuigan (Ed.) *The Psychophysiology of Thinking*. New York: Academic Press, 1973.

Megaw, E. D. Direction and extent uncertainty in step tracking. *Journal of Motor Behavior*, 1972, *4*, 171–186.

Russell, D. G. Spatial location cues and movement production. In G. E. Stelmach (Ed.), *Motor control: Issues and trends*. New York: Academic Press, 1976.

Sanders, A. F. Some variables affecting the relation between relative stimulus frequency and choice reaction time. *Acta Psychologica*, 1970, *33*, 45–55.

Schmidt, R. A. The index of preprogramming (IP): A statistical method for evaluating the role of feedback in simple movements. *Psychonomic Science*, 1972, *27*, 83–85.

Seibel, R. Data entry devices and procedures. In H. P. Van Cott and R. G. Kinkade (Eds.) *Human engineering guide to equipment design*. Washington, D.C.: American Institute for Research, 1972.

Shaffer, L. H. Intention and performance. *Psychological Review*, 1976, *83*, 375–393.

Shaffer, L. H., & Hardwick, J. Errors and error detection in typing. *Quarterly Journal of Experimental Psychology,* 1969, *21,* 209–213.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica,* 1969, *30,* 276–315.(a)

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist,* 1969, *57,* 421–457.(b)

Van Nes, F. L. Analysis of keying errors. *Ergonomics,* 1976, *19,* 165–174.

Welford, A. T. On the sequencing of action. *Brain Research,* 1974, *71,* 381–392.

Wickelgren, W. A. Context-sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review,* 1969, *76,* 1–15.